

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

I-Lin Wu

Date

Multiple site-specific incorporation of non-canonical amino acids for novel biomaterials design

By

I-Lin Wu
Doctor of Philosophy

Department of Chemistry

Vincent P. Conticello, Ph.D.
Adviser

Stefan Lutz, Ph.D.
Committee Member

Khalid Salaita, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Multiple site-specific incorporation of non-canonical amino acids for novel biomaterials design

By

I-Lin Wu

B.S., National Taiwan University, 2004

M.S., National Taiwan University, 2006

Adviser

Vincent P. Conticello, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Chemistry

2013

Abstract

Multiple site-specific incorporation of non-canonical amino acids for novel biomaterials design

By I-Lin Wu

Site-specific incorporation of selective unnatural functional groups into proteins has provided a means to expand the protein diversity, creating proteins with unique chemical properties useful for a broad range of new applications. However, the preparation of protein polymers derived from the sequence-repetitive polypeptides requires new development in which non-canonical amino acids are incorporated at multiple and specific locations in the polypeptides sequences. We describe herein a simple and efficient method to facilitate the multi-site selective insertions of non-canonical amino acids at structurally defined positions within recombinant polypeptides and provide opportunities for engineering proteins modified extensively with selected amino acid analogues. In this approach, *Escherichia coli* MRA30, a bacterial host strain with an attenuated activity of release factor 1 (RF1), was assessed for its ability to support the incorporation of a diverse range of non-canonical amino acids in response to multiple encoded amber codons (UAG) within genes derived from superfolder GFP and elastin-mimetic protein polymers. Suppression efficiency and protein yield depended on the identity of the orthogonal aminoacyl-tRNA synthetase/ tRNA_{CUA} pair and the non-canonical amino acid. Elastin-mimetic protein polymers were prepared where non-canonical amino acids were incorporated up to twenty-two specific sites with high substitution efficiency. The identities and positions of the variant residues were confirmed by mass spectrometric analysis of the full-length polypeptides and the proteolytic cleavage fragments from the thermolysin digestion. Based on the developed system, we further demonstrate the generation of novel amphiphilic elastin diblock copolymers that could undergo temperature-dependent segregation and self-assemble into core-shell-corona nanoparticles with photo-crosslinkable activities. The photo-crosslinking experiments were monitored by dynamic light scattering and microscopes. Our data suggest that this multi-site suppression approach permits the preparation of protein-based materials in which novel chemical functionality can be introduced at precisely pre-defined positions within the polypeptide sequence.

Multiple site-specific incorporation of non-canonical amino acids for novel biomaterials design

By

I-Lin Wu

B.S., National Taiwan University, 2004

M.S., National Taiwan University, 2006

Adviser

Vincent P. Conticello, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2013

Acknowledgements

Ph. D is a long and winding road; you never know what might happen until you reach to the end. Once you are there and look back, realized the long journeys you've traveled, you will truly feel grateful for the help and love you got when you are en route there. There have been a lot of up-and-downs not only in science but also in life on the road. Without the great help from a lot people, I shall not be able to accomplish and defend my thesis. I want to express my deepest gratitude to my advisor Prof. Vince Conticello. With his guidance and mentorship, I could always stay on the right track to the goal. Mentoring a student for six years may not be harder than raising a child, but it is still somewhat difficult to work with an emotional student and calm her down when life is not going, as it desires. I would also like to thank Prof. Stefan Lutz and Prof. Khalid Salaita for their supervision during my graduate study. Their enthusiasms in scientific research have inspired me a lot in developing scientific arguments. I have to thank Prof. Chaikof for stimulating ideas on the elastin research projects. I also want to thank Dr. Fred Strobel for having me work in the mass spectrometry room during these years. I have learned many techniques and histories of mass spectrometry. As my life changing moment, I want to thank Prof. Gianluca for the opportunity to shadow him in Italy with beautiful and memorable scenery. The life and research experience in Italy was wonderful and true to be a landmark in my ph. D journey.

I really enjoyed working in the "Conticello Lab" as our lab is the most people-friendly working environment for daily life. Holly taught me a lot when I first came and constantly supports me afterwards. Yunyun and Melissa, we shared years together for laughs and cries. Chris and Tao, are my spiritual supports and research dictionaries. My

great lab members, Rui, Paolo, Elizabeth, Becky, Charlie, Spencer, are comforting sweethearts when I was out of control. My loyal coffee-mates, Weilin and Yihan, are necessities of the ph. D. I greatly appreciate their persistent friendship and lifelong support in science and life. For a lot of friends, we met in Atlanta, and experienced the fun-and-down together, Fan, Jenny, Yun, Yuhwa, thank for the company and the help for these years. Finally, I would like to thank my best friends in Taiwan, Shevonne, Miin, Anne, Sheila, and especially Ibiza and Angex. Without your help, I never image myself studying aboard alone. My lovely cat, Mini, is a loyal company always and all the time.

Last but not least, I would love to give my greatest appreciation to my dad and mom. They are always there for help with no reasons and complaints. Life has been changed a lot during this journey, the only thing that remained constant, is the forever love from them.

Dedicated to my Mom and Dad

I-Lin

Table of Contents

Multiple site-specific incorporation of non-canonical amino acids for novel biomaterials design

Chapter 1. Introduction	1
1. Incorporation of non-canonical amino acids	2
2. Protein-based biomaterials	6
3. Elastin	11
4. Elastin-mimetic polypeptides	14
5. Amphiphilic block copolymers	16
6. References	22
Chapter 2. System development for the multi-site specific incorporation of non-canonical amino acids	25
1. Introduction	26
2. Experimental Methods	38
a. Materials	38
b. General methods	39
c. Construction of the expression vector	40
d. Construction of sfGFP expression plasmids	44
e. Construction of the elastin-mimetic peptides, Elastin-UAG genes	45
f. Construction of the Elastin-UAG expression plasmids	48
g. Construction of the orthogonal aminoacyl-tRNA synthetase/tRNA pairs	49
h. Protein expression and purification	51
i. Flow Cytometry	55
j. Thermolysin digestion	56
k. Mass spectrometry	57
l. Temperature-dependent turbidity	58

3. Results and Discussion	68
4. Conclusions	90
5. References	92
Chapter 3. System optimization for the multi-site specific incorporation of photo-crosslinkable amino acid analogues	96
1. Introduction	97
2. Experimental Methods:	104
a. Materials	104
b. General methods	105
c. Construction of the <i>M. jannaschii</i> tRNA plasmid	106
d. Construction of the <i>M. jannaschii</i> aminoacyl-tRNA synthetase / tRNA plasmid	106
e. Bacterial growth and expression	107
f. TALON [®] metal-affinity column purification	108
g. Flow cytometry	110
h. Thermolysin digestion	110
i. Mass spectrometry	111
j. Photo-crosslink experiments	112
3. Results and discussion	116
4. Conclusions	138
5. References	139
Chapter 4. System application: Synthesis of photo-crosslinkable elastin diblock copolymers	143
1. Introduction	144
2. Experimental Methods:	152
a. Materials	152
b. General methods	152
a. Construction of the Elastin-diblock plasmids	153

b. Protein expression and purification	155
c. Transmission electron microscopy	156
d. Temperature-dependent turbidity	157
e. Dynamic light scattering	157
f. Atomic force microscopy	158
g. Thermolysin digestion and Mass spectrometry	158
3. Results and discussion	164
4. Conclusions	178
5. References	180
Chapter 5. Conclusion	184
1. Conclusion	185
2. References	191
Appendix. Sequences of Interest	193
Appendix 1. Sequencing primers utilized in the study	194
Appendix 2. Sequences of the MbPylRS	195
Appendix 3. Sequences of the MjTyrRS	196
Appendix 4. Sequences of the wild-type sfGFP-CCC	197
Appendix 5. Sequences of the wild-type sfGFP-(UAG) ₄	198

List of Figures

Multiple site-specific incorporation of non-canonical amino acids for novel biomaterials design

Chapter 1.

Figure 1. Site-specific incorporation of UAA via orthogonal aminoacyl-tRNA and tRNA pair	5
Figure 2. Selection schemes for genetically encoding amino acids in <i>E. coli</i> and <i>yeast</i>	7

Figure 3. Chemical structures of genetically encoded unnatural amino acids	8
Figure 4. Natural proteins for bio-inspired materials	9
Figure 5. Multifunctional protein-based block copolymers	11
Figure 6. Inverse temperature transition (T_i) of poly[f_v (VPGVG), f_x (VPGXG)]	15
Figure 7. Inverse temperature transition (T_i) depends on the chain length for a series of three ELPs based on (Val-Pro-Gly-Xaa-Gly) repeat units	16
Chapter 2.	
Figure 1. Synthesis of repetitive polypeptides via concatemerizaion of DNA cassettes	28
Figure 2. Crystal structures of <i>M. mazei</i> pyrrolysine synthetase (PylRS)	31
Figure 3. Structure-based sequence alignments of <i>M. mazei</i> PylRS with other PylRSs.	32
Figure 4. Secondary structure of the sfGFP variant depicting the sites of UAG codon insertion	34
Figure 5. Plasmid map of PQE-80L	41
Figure 6. Modified pQE-80L plasmids with adaptor sequence	44
Figure 7. Plasmid map of sfGFP expression plasmids	45
Figure 8. Plasmid map of pZErO [®] -1	47
Figure 9. Construction of the Elastin-UAG multimers	59
Figure 10. Construction of Elastin-UAG expression vectors	60
Figure 11. Plasmid map of pSU81 and pHEC2	61
Figure 12. Plasmid pIL38 and pIL39 encoding MbtRNA _{CUA} in pSU81 and pHEC2 respectively	62
Figure 13. Plasmids encode MbPylRS pairs in pHEC2	63
Figure 14. FACS analysis of sfGFP derivitaves for the indicated non-canonical amino acid	71
Figure 15. Expression studies of sfGFP derivatives modified with pyrrolysine analogues	73
Figure 16. Deconvoluted ESI-mass spectra of purified sfGFP derivatives	74

Figure 17. Expression of elastin-(UAG) _n derivatives with pyrrolysine analogues	79
Figure 18. Temperature-dependent turbidimetry profiles for elastin derivatives	81
Figure 19. MALDI-TOF mass spectra of full-length elastin derivatives	85
Figure 20. ESI-MS and MS/MS spectra of elastin-(AlocLys) ₁₂ derivatives digested with thermolysin	87
Figure 21. ESI-MS and MS/MS spectra of elastin-(tBocLys) ₁₂ derivatives digested with thermolysin	88
Figure 22. ESI-MS and MS/MS spectra of elastin-(CbzLys) ₁₂ derivatives digested with thermolysin	89
Chapter 3.	
Figure 1. Bpa and AzF serve as photo-crosslinkers	100
Figure 2. The EF-Tu/ tRNA interface	102
Figure 3. Construction of evolved MjtRNA _{CUA} expression plasmids	118
Figure 4. Expression plasmids of MjTyrRS and MjBpaRS with evolved MjtRNA _{CUA} .	119
Figure 5. FACS analysis of sfGFP-Bpa for the indicated BpaRS pair	120
Figure 6. Western blot analysis for the expression of sfGFP-(UAG) ₃ with BpaRS pairs in <i>E.coli</i> MRA30 and DH10B strain	121
Figure 7. Expression plasmids of MjAzFRS with evolved MjtRNA _{CUA} .	122
Figure 8. FACS analysis of sfGFP-AzF for the indicated AzFRS pair	123
Figure 9. Western blot analysis of the expression of sfGFP-(UAG) ₃ with AzFRS pair	125
Figure 10. Deconvoluted ESI-mass spectra of purified sfGFP derivatives	126
Figure 11. Expression and purification of elastin-(UAG) _n derivatives	131
Figure 12. MALDI-TOF spectra of the full-length elastin derivative, elastin-(Bpa) ₁₂ and elastin-(AzF) ₁₂	132
Figure 13. ESI-MS and MS/MS spectra of elastin-(Bpa) ₁₂ derivatives digested with thermolysin	133

Figure 14. MALDI-TOF and TOF/TOF spectra and data analysis of digested peptide fragment, [VGVPG(Bpa)GVPG] ⁺	134
Figure 15. ESI-MS and MS/MS spectra of elastin-(AzF) ₁₂ derivatives digested with thermolysin	135
Figure 16. Gel electrophoresis of sfGFP-(AzF) ₃ after UV-induced cross-linking	137
Figure 17. Gel electrophoresis of elastin-(AzF) ₁₂ after UV-induced cross-linking	137
Chapter 4.	
Figure 1. Schematic of expected behaviors of elastin diblock copolymers	146
Figure 2. Disulfide cross-linked elastin diblock copolymers	147
Figure 3. Directional cloning strategy for assembly of synthetic genes encoding diblock polypeptides	160
Figure 4. Plasmid map of pZErO [®] -2	161
Figure 5. Plasmids pSCP3.11 and pIL103.12 encoding elastin-S1 and elastin-UAG concatemers in pZErO [®] -2. Plasmid pIL42.12 encoding elastin-UAG concatemers in pZero [®] -1	162
Figure 6. Plasmid pIL105 encoding diblock gene fusion in pZErO [®] -2. Plasmid pIL106 encoding the diblock gene fusion in the expression plasmid pIL5	163
Figure 7. Expression and purification of elastin diblock copolymers	166
Figure 8. Temperature-dependent turbidimetry profiles of elastin-(AzF) ₁₂ and ELP-AzF diblock copolymers	169
Figure 9. TEM images of ELP-AzF copolymers at various pHs and temperatures	171
Figure 10. TEM images of the self-assemble behaviors of the ELP-AzF diblock copolymers with or without UV crosslinking	172
Figure 11. AFM images of the photo-crosslinked ELP-AzF diblock copolymers	173
Figure 12. DLS of the ELP-AzF diblock copolymers	175
Figure 13. Gel electrophoresis of ELP-AzF diblock copolymer after UV-induced cross-linking	176

Figure 14. MALDI-TOF spectra of thermolysin digested non-crosslinked and UV-crosslinked ELP-AzF diblock copolymers	177
---	-----

Chapter 5. Conclusion

Figure 1. Multi-site specific incorporation of non-canonical amino acids in <i>E. coli</i>	190
---	-----

List of Tables.

Chapter 2.

Table 1. Plasmids and strains utilized in Chapter 2	64
Table 2. Primers utilized in Chapter 2	66
Table 3. Superfolder GFP expression and ESI-mass spectrometry data	75
Table 4. Elastin protein expression and MALDI-TOF mass spectrometry data	86

Chapter 3.

Table 1. Plasmid and strains used in Chapter 3	113
Table 2. Primers used in Chapter 3	115
Table 3. Superfolder GFP expression and ESI-mass spectrometry data	127
Table 4. Elastin protein expression and MALDI-TOF mass spectrometry data	130

Chapter 4.

Table 1. Plasmids and strains used in Chapter 4	159
--	-----

List of Schemes

Chapter 2.

Scheme 1. Chemical structures of pyrrolysine analogues.	33
Scheme 2. Oligonucleotide cassettes encoding the elastin-UAG monomer sequences	37
Scheme 3. Sequence of pQE-Adaptor for cloning of expression plasmids	43

Chapter 1

Introduction

Introduction

Incorporation of non-canonical amino acids

Proteins depend upon post-translational modifications for their proper functioning in the cell¹; however, only twenty natural amino acids are encoded by the defined genetic code system. While site-directed mutagenesis allows scientists to modify protein composition and function, the chemical functionality among natural amino acids provides a relatively limited selection. The ability to incorporate amino acid analogues with properties that differ from their native counterparts provides a powerful tool for exploring protein structure and function *in vitro* and *in vivo*. These unique functional groups could be employed as chemical and physical probes to investigate the role of specific amino acid residues on protein biochemistry, including their influence on protein folding and thermodynamic stability, substrate recognition, enzymatic catalysis, and protein-based materials with expanded technological utility². As a result, unnatural amino acids with an array range of structural and electronic properties have extended our ability to manipulate the physicochemical and biological properties of engineered proteins.

Methods based on chemical modification³, solid-phase synthesis⁴ and expressed protein ligation⁵, have been demonstrated to introduce reactive functional groups, biophysical probes, and other non-peptidic backbones into peptides and proteins⁶. These strategies are limited by the homogeneity of the synthesized protein and also the number of residues that can be modified. Alternatively, the residue-specific method allows the global substitution of an amino acid with a close structural analogue throughout the proteome, typically in competition with the corresponding amino acid^{7,8}. This approach 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^{8,9}. Similarity between the analogues and the canonical amino acid allows them to be transformed into cells and charged by appropriate aaRS. Generally, protein expression is carried out in a strain auxotrophic for a specific amino acid and starved of the natural amino acid but supplemented with the analogue in chemically defined media. Although residue-specific incorporation is a relative simple method that allows for substantial alterations in the physical properties of proteins and enables chemical modification at multiple sites, this approach leads to a proteome-wide replacement bearing unnatural amino acids in the place of every occurrence of a particular natural amino acid, rather than the quantitative installation of an unnatural amino acid at a defined site in a protein. In addition, the use of these strategies is limited by selection of protein sequence and toleration of wild-type¹⁰ or mutated cognate aaRS¹¹ toward amino acid analogues. 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 aaRS. 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.

Site-specific incorporation of selective unnatural functional groups within proteins provides a controllable means to expand protein diversity, creating proteins with unique chemical properties¹²⁻¹⁴. Biosynthetic approaches have also been developed for the *in vitro* and *in vivo* synthesis of proteins containing non-canonical amino acids site-specifically. *In vitro* biosynthetic routes translated in cell-free systems usually involve tRNAs that recognize nonsense, frameshift, or coding codons, which are chemically

aminoacylated with the unnatural amino acid target¹⁵⁻¹⁸. This method has been used for incorporation of amino acids with modified backbones¹⁹, functional groups related to post-translational modifications²⁰, and photo- and chemically reactive side chains²¹. In addition, unnatural amino acids spanning a range of electronic, structural, and conformational properties have been incorporated into proteins by microinjection or transfection of the aminoacylated tRNAs into living cells²²⁻²⁴. However, 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 viability of cells by the microinjection and transfection techniques.

Genetic engineering is powerful for protein modification and *de novo* design of artificial proteins. Schultz and co-workers have demonstrated effective methods for incorporating unnatural amino acids and analogues using unique triplet or quadruplet codons^{25,26}. These earlier approaches took advantage of the termination codons UAA, UAG, and UGA that do not encode amino acids. These “blank” codons that exist in the genetic code can be used to uniquely specify a non-cognate amino acid or analogue with suppressor tRNA. In general terms, methods to expand the genetic code have included the selection and directed evolution of orthogonal aaRS/tRNA pairs that function in a heterologous host²⁷. This approach is based on the observation that some prokaryotic aaRS/tRNA pairs do not cross-react to any significant degree with their eukaryotic or archaeal counterparts and function independently in the host organism. To ensure the success of this strategy, the components of the expression system must satisfy a number of criteria (Fig. 1). First, the orthogonal tRNA/synthetase pair should recognize the corresponding unnatural amino acid. Second, the blank codon must be recognized by the

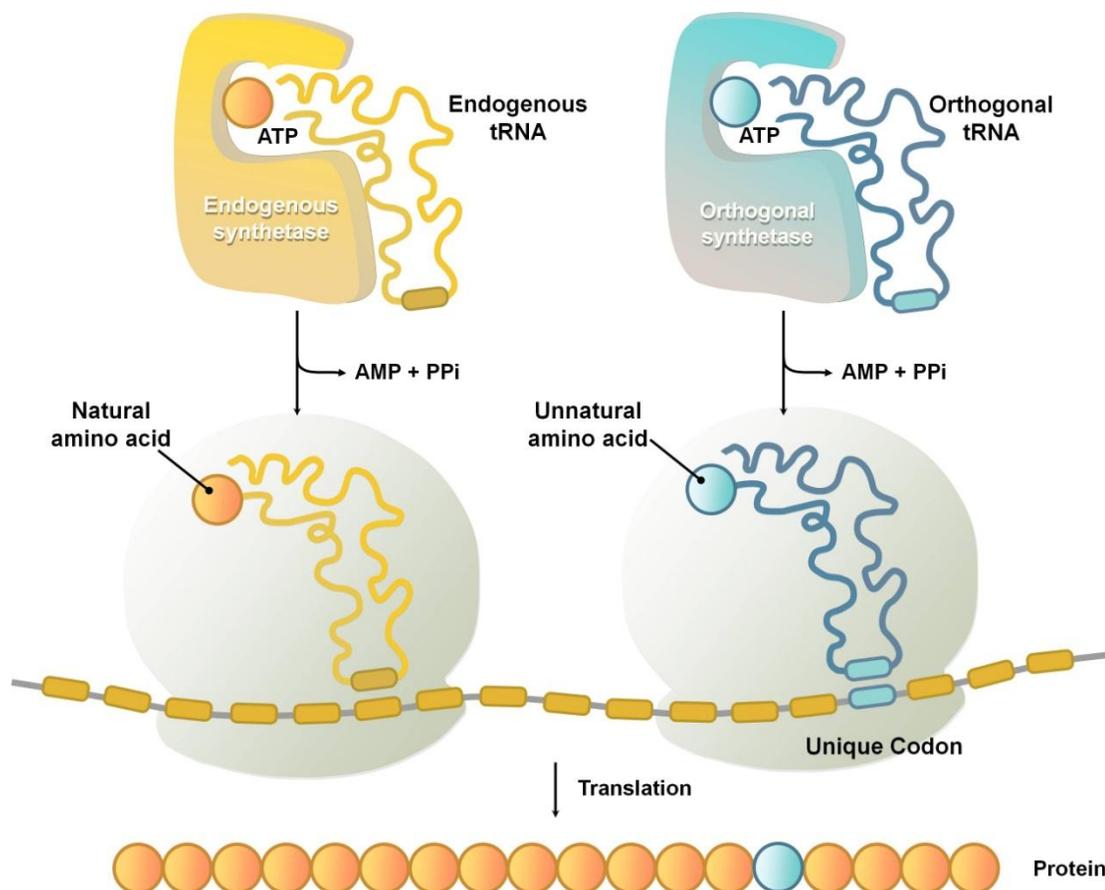


Figure 1. Site-specific incorporation of UAA via orthogonal aminoacyl-tRNA and tRNA pair. The new components for genetic code expansion consist of an orthogonal tRNA, and orthogonal aminoacyl-tRNA synthetase, and a unique codon. The orthogonal tRNA is acylated with the UAA specifically by the orthogonal synthetase. During protein translation, the acylated tRNA delivers the UAA in response to a "unique codon" that does not encode for any of the canonical amino acids.

orthogonal tRNA, but not any endogenous tRNAs. Third, the orthogonal tRNA/synthetase pair must be specific for the unnatural amino acid, compatible with the elongation factor and ribosome, and exhibit no cross-reactivity with any endogenous tRNA/synthetase pairs in the host organism. Schultz and coworkers have made important contributions to the development of this strategy²⁸. High-throughput selection schemes based on cell survival assay enable the evolution of aminoacyl-tRNA synthetases that show high specificity toward structurally dissimilar unnatural amino acid analogues with technologically useful functional group (Fig. 2)¹². Currently, over 70 novel amino acids, such as fluorescent²⁹, photo-reactive³⁰, NMR probes³¹, and other amino acid analogues have been genetically encoded in response to termination codons in bacteria, yeast and mammalian cells by selected and evolved aaRS pairs (Fig. 3)¹². However, due to competition with release factors, this scheme needs to be optimized for multiple-site selective incorporation of unnatural amino acid analogues. In addition, the development of an efficient screening system to assay the multiple suppression events of termination codons is highly desirable.

Protein-based biomaterials

Natural structural proteins play critical structural and mechanical roles in biological systems and have distinguished mechanical, chemical, electrical, and optical properties. Most structural proteins such as elastins, collagens, silks, and resilins are of tremendous interest as targets for protein-based materials design^{32,33}. They are characterized with a long-range ordered molecular secondary structures such as β -sheets, coiled-coils, or triple helices that spread within the proteins due to their highly repetitive

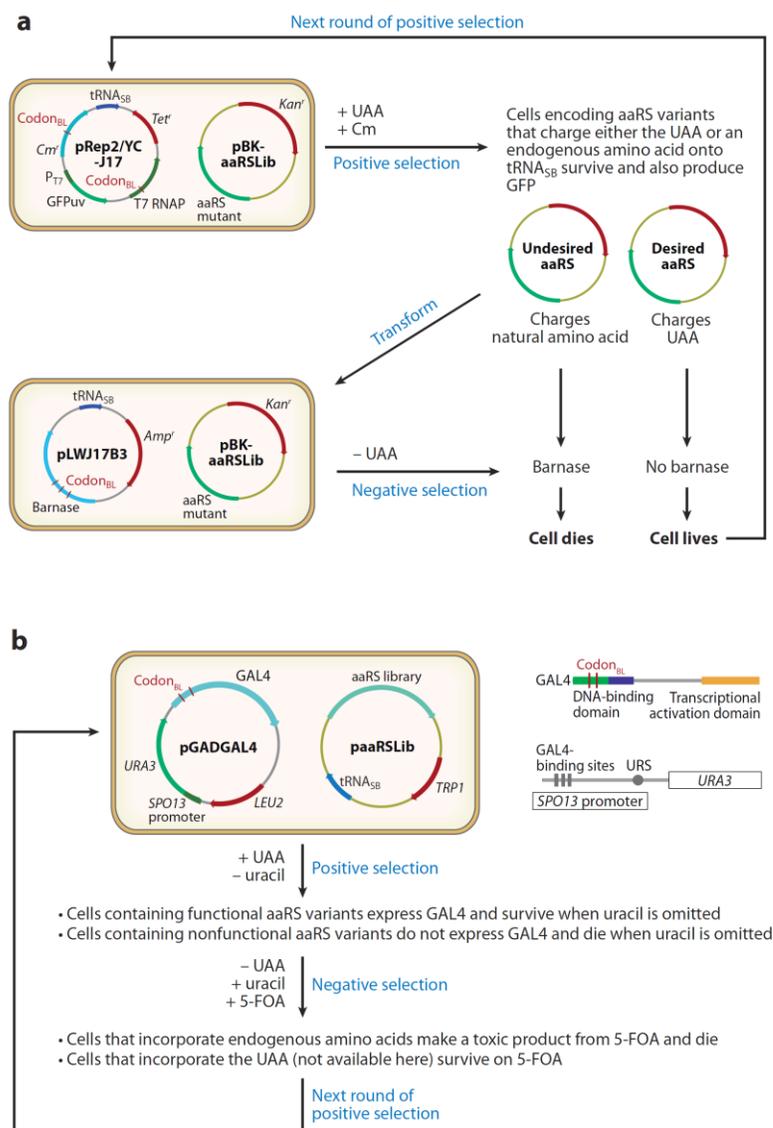


Figure 2. Selection schemes for genetically encoding amino acids in *E. coli* and yeast.¹²

Two-step selection for the directed evolution of orthogonal UAA-RS/tRNA pairs in (a) *E. coli*. (b) yeast.

Abbreviations: aaRS, aminoacyl-tRNA synthetase; Condon_{BL}, blank codon; 5-FOA, 5-fluoroorotic acid; UAA, unnatural amino acid.

sequences (Fig. 4)³³. These specific features promote their self-assembly to form a structural hierarchy with distinct mechanical functions in nature. The primary structure of these native materials presumably determines their unique structural and functional properties.³⁴ In addition, the sequence specificity strongly influences the mechanism of self-assembly of the polymer chains into the supramolecular architectures that characterize the materials properties and performances. Therefore, these defined motifs

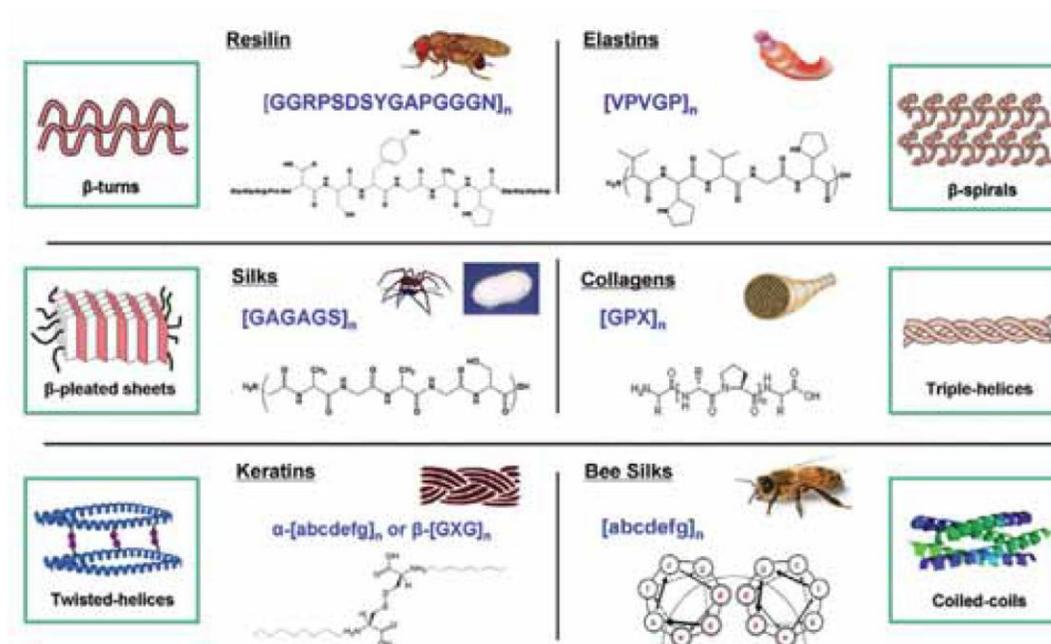


Figure 4. Natural proteins for bio-inspired materials. Natural structural proteins including resilins, elastins, spider and silkworm silks, collagens, hair keratins, and bee silks³³.

derived from the native structural proteins have been employed as a "mimetic core" to represent their bioactive functionalities and therefore envisioned in biomaterials applications.

Genetic engineering has enabled the creation of protein materials derived from sequence-repetitive polypeptides with defined physical, chemical, and biological properties. The modular design strategy employed in biomaterials engineering allows for

the inclusion of multiple peptide domains with distinct functionalities into one polypeptide³⁵⁻³⁷. For example, combinations of different protein blocks could generate new multifunctional protein composite systems with tailorable properties for specific mechanical, electrical, optical, or thermal applications (Fig. 5)³³. These block copolymers consist of well-defined blocks where the individual blocks display different mechanical, chemical, or biological properties. The controlled architectural uniformity of the macromolecules promotes self-assembly into structurally defined supra-molecular materials that exhibit distinctive properties as a consequence of this sequence specificity^{38,39}. For example, a range of mechanical properties can be achieved to mimic the range of elasticities of different tissues by mixing rigid and soft protein such as silk and tropoelastin in different ratios^{40,41}. In addition, self-assembly of polypeptides has led to a number of materials with precise structural control at molecular levels, such as films^{42,43}, fibers^{44,45}, micelles⁴⁶⁻⁴⁹, and hydrogels⁵⁰⁻⁵³ that have tremendous potential in drug delivery and regenerative medicine. These provide a tool-kit of biopolymer options with different material properties, suggesting that blending blocks is an efficient approach to generate biomaterials with a broader range of properties or with properties that capture a combination of features from the components. Recombinant protein expression allows for the synthesis of proteins with the functionality of naturally derived materials, while overcoming their limitations through highly reproducible and modular chemical compositions, making them ideal for many biomedical applications over conventional synthetic polymers⁵⁴. However, ribosomal protein synthesis affords polypeptide sequences with limited chemical diversity, particularly in comparison to organic polymers derived from chemical synthesis. The development of methods that facilitate

the multi-site substitution of amino acid analogues represents a significant advance in biosynthetic technology, promising access to protein-based materials with multiple non-canonical amino acids encoded at defined sites within the polypeptide sequence⁵⁵. The

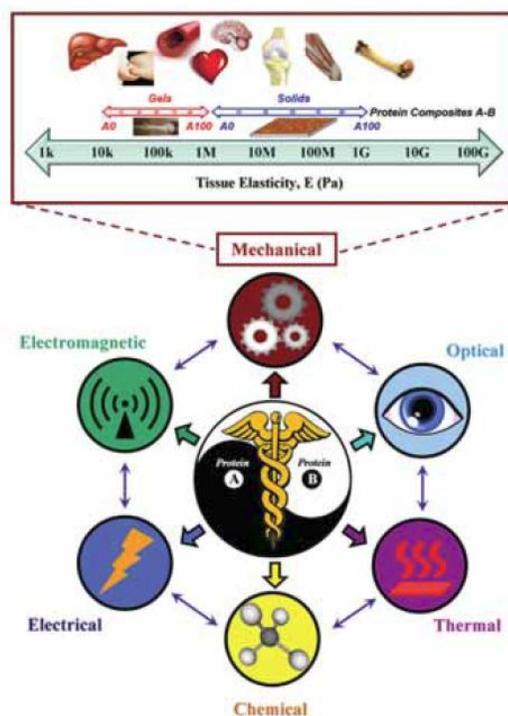


Figure 5. Multifunctional protein-based block copolymers. Combination of protein A and B can generate new multifunctional protein composite systems with controllable properties³³.

synthetic approach described in this study permits the incorporation of novel chemical groups into protein-based materials that may be employed to expand the range of functional properties with respect to the native protein-based materials.

Elastin

Elastin is a major component of extracellular matrix proteins widely distributed in vertebrate tissues that imparts extensibility and resiliency to tissues such as the dermis,

lung, and larger blood vessels. Elastin networks confer elasticity to the connective tissues, undergoing high deformation without rupture and recovering to its original state once stress is removed⁵⁶. Elastin is characterized with two significant structural features inherent in all elastomeric materials. First, it is composed of monomers consisting of repetitive glycine-rich peptide motifs that are flexible and conformationally unconstrained, enabling an area to exhibit reversible deformation. Secondly, the size and properties of the mobile monomer domain and the degree of intermolecular cross-linking can influence the elastic behavior of the resulting elastin network.

Elastin is formed from the soluble precursor protein tropoelastin. Tropoelastin consists of alternating hydrophobic domains rich in non-polar amino acids and hydrophilic domains rich in lysine residues separated by alanine residues⁵⁷. The hydrophobic domains are characterized by three to six peptide repeats with sequences such as GVGVP, GGVP, and GVGVP important for elastic properties. The hydrophilic domains are rich in polyalanyl-co-lysine stretches involved in cross-linking. The hydrophilic domain displays α -helical character while the hydrophobic domains exist in a tandem coil conformation with the presence of type II like β -turns within the structure. Monomers of tropoelastin are covalently cross-linked to form a loosely ordered three-dimensional fibrous network with lysine residues that are dispersed throughout the protein. Elastin fibers appear to exist as two morphologically different components *in vivo*, a highly isotropic amorphous elastin constituent within an organized microfibrillar, cross-linked scaffold primarily composed of fibrillin⁵⁷. The coacervation mechanism of soluble tropoelastin caused by interaction between the hydrophobic domains is believed to be important for fiber formation⁵⁸. Increasing the temperature of soluble elastin above

the inverse temperature transition (T_t) results in a turbid solution due to the self-assembly and aggregation of the hydrophobic chains of elastin molecules that form a coacervate. The association of elastin molecules is thermodynamically controlled, and is readily reversible as the coacervate redissolves upon cooling of the solution. The coacervate phase is responsible for the viscoelastic properties of native elastic fiber that are integral to its biological function^{59,60}.

Native elastin network is formed through the enzymatic cross-linking of tropoelastin into an insoluble matrix upon deposition. The cross-linking regions consisted of pairs of lysine residues separated by alanine (KA) or proline (KP). The hydrophobic domain was found to be composed of up to eleven repeats of VPGVG. After secretion, inter- and intra-molecular cross-links are formed through the deamination of the ϵ -amino group of the lysine side-chains by the copper-requiring enzyme lysyl oxidase⁶¹. The reactive aldehyde condenses with a second aldehyde residue to form allysine aldol or with an ϵ -amino group on lysine to form dehydrolysinovaline. These two precursors could further condense to form the unique tetrafunctional cross-links such as desmosine and isodesmosine⁶². The composition and sequence of these domains strongly influence the local polypeptide chain conformation and determine the positions and density of the cross-links. The structural uniformity of these natural elastic materials is presumably responsible for their unique elastomeric mechanical responses or resilience properties. The ability of elastin to undergo reversible deformation, stretch under large amount of stress, form hydrogel matrices, and undergo an inverse temperature transition make it an attractive material to emulate for the design of tissue-engineered analogues with potential biomedical applications. The well-defined correlation between elastin sequence and

macromolecular properties enables the construction of a wide range of biosynthetic analogues with tunable biophysical properties^{63,64}. The biosynthesis of native elastin networks could serve as a paradigm for the design of protein-based materials to retain the structural uniformity of their natural counterparts, while permitting the introduction of functional groups that facilitate formation of extended networks under controllable conditions.

Elastin-mimetic polypeptides

Elastin-mimetic polypeptides have been developed as soluble mimics of native elastin. Consensus tetra-(VPGG), penta-(VPGVG), and hexa-(APGVGV) peptide motifs have been used in biomaterials to mimic the characteristic properties of native elastin⁶⁵⁻⁶⁷. Inverse temperature transition (T_i) is an appealing property of elastin-mimetic polypeptides for use as drug delivery systems and other medical applications. The assembly of the elastin-mimetic polymer above the T_i is the result of spontaneous phase separation of the polypeptide, which coincides with a conformational rearrangement of the local secondary structure from a soluble disordered chain to a collapsed insoluble aggregate. Urry and coworkers have made extensive studies using synthetic elastins with the repeating sequence [(Val-Pro-Gly-Xaa-Gly)]^{63,64}. Polypeptides based on the pentameric repeat sequences undergo a reversible, temperature-dependent, hydrophobic assembly from aqueous solution to the formation of a coacervate. The presence of the guest residues, which is usually at the fourth position (Xaa) within the pentapeptide unit, has been employed as a mechanism to alter its temperature-dependent phase behavior and mechanical properties⁵¹. It was shown that the Xaa position tolerates a wide variety of non-conservative amino acid substitutions, which alters the position of the inverse

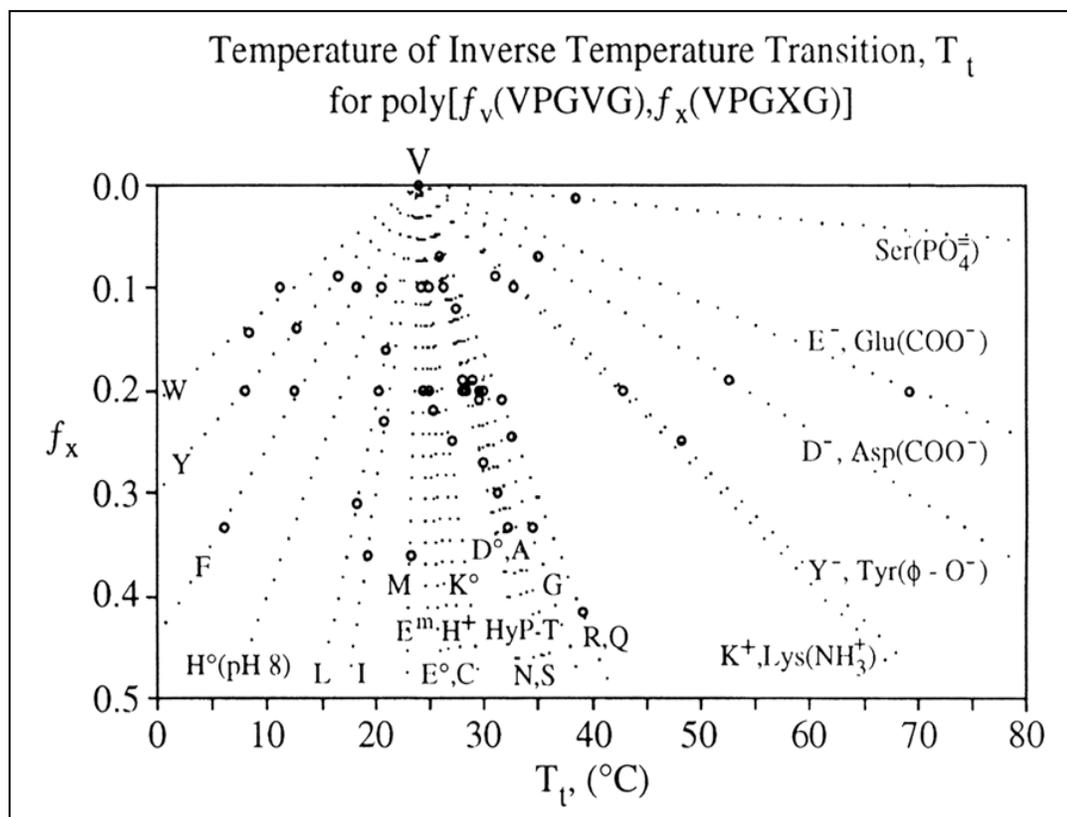


Figure 6. Inverse temperature transition (T_t) of poly[f_v (VPGVG), f_x (VPGXG)]. Hydrophobic scale based on T_t values for the copolymers described above as a function of guest residue identity and mole fraction, f_x . Hydrophobic amino acids shift the position of T_t toward lower temperature and more polar residues shift the position of T_t toward higher temperatures in a manner commensurate with the level of substitution within the polypeptide sequence.⁶³

temperature transition (Fig. 6)⁶³. The T_t can be raised or lowered on the basis of the hydrophobicity profile of the incorporated amino acids. In later studies, Chilkoti and coworkers demonstrated that, for a given elastin-mimetic repeat sequence, the coacervation temperature behaved as a function of the length of polypeptide repeats (Fig. 7)⁶⁸. Furthermore, replacement with Ala for the consensus Gly residue in third position of the pentapeptide repeat induces a change in the mechanical response of the material from

elastic to plastic. The effect has been attributed to a change from a type II to a type I β -turn arrangement at Pro-Gly. Thus, the elastin-mimetic polypeptides based on the repeat

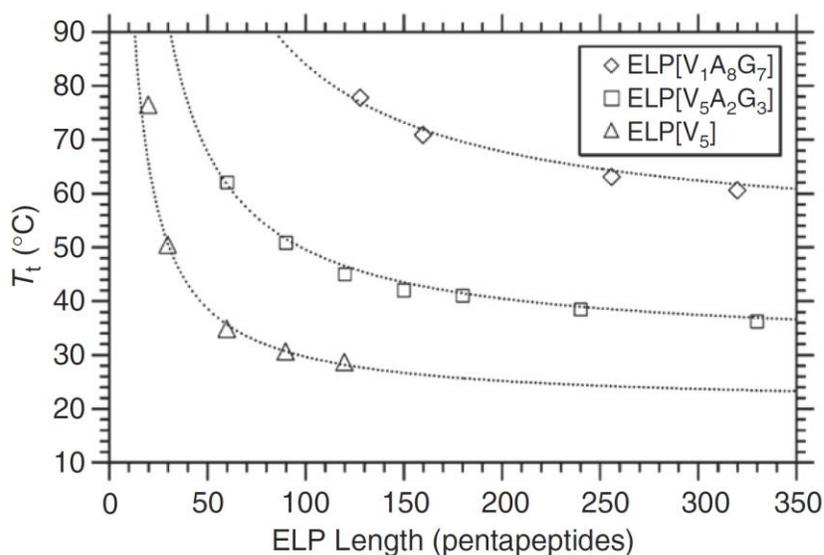


Figure 7. Inverse temperature transition (T_t) depends on the chain length for a series of three ELPs based on (Val-Pro-Gly-Xaa-Gly) repeat units. The letters within the brackets indicate the identity and relative number of the guest residues, Xaa, within the pentapeptide repeat units of the respective ELPs.⁶⁸

sequence [(Val-Pro-Gly-Xaa-Gly)] display a well-defined correlation between the repeat sequence and macromolecular properties, which is a suitable peptide sequence to use as a building block for the creation of a synthetic elastin-like protein (ELP).

Amphiphilic Block Copolymers

Block copolymers are a type of polymer in which two or more well-defined blocks where the individual blocks display different mechanical, chemical, or biological properties are covalently linked. Many conventional block copolymers of technological

significance exhibit amphiphilicity. Generally, amphiphilic block copolymers are composed of compositionally distinct blocks that exhibit distinctly different interaction affinity for aqueous solutions. Amphiphilic diblock (**AB**) and triblock (**ABA** and **BAB**) copolymers are formed by two incompatible polymer domain that under selective segregation of the hydrophobic domain in aqueous solvents to form well-defined mesoscopic aggregates. In the diblock copolymer, the microdomains consist of a dense core of hydrophobic block from which a corona of flexible hydrophilic block chains emerge and reveal a spherical, rod, thread, or disk-like shape. Elastin-mimetic blocks vary based on the primary sequence design to provide versatility in the mechanical properties or thermo-responsive activities of the synthetic construct. The genetic engineering of elastin-like proteins (ELPs) permits the formation of block copolymers from individual block polymer. The phase separation of hydrophobic domains within ELP block copolymers can be exploited as a general mechanism for the reversible self-assembly of protein-based materials. Previously, our lab has reported the synthesis of the elastin triblock copolymers (**BAB**-type) where the hydrophobic block (**B**) is derived from the plastic pentapeptide sequence [Val/Ile-Pro-Ala-Val-Gly] with T_t below 37 °C and the hydrophilic block (**A**) based on the elastomeric repeat sequence [Val-Pro-Gly-Glu-Gly] display higher T_t values⁵². The resulting structure consist of a solvent-swollen network of spherical particles (micelles) derived from the insoluble endblock above transition and linked together through the soluble midblock. The hydrophobic protein chains will entangle with the solvent swollen domain to form a physically cross-linked hydrogel elastin network^{69,70}. In addition, a glutaraldehyde cross-linking method has been applied to the similar elastin triblock copolymer interspersed with lysine-rich domains permitting

both physical and covalent crosslinking. Subsequent studies showed that the length ratio of un-cross-linked block versus cross-linked block and the block copolymer architecture had a significant effect on the swelling ratio of the cross-linked hydrogels, their microstructure, and mechanical properties. Moreover, enhanced cell viability within triblock ELP hydrogels was also observed to indicate the biocompatibility of ELP block copolymers in biomedical application⁷⁰. The phase behavior of ELP amphiphilic block copolymers can be modified through manipulation of the length, composition, and sequence of the individual blocks and thus self-assemble into a variety of diverse structures including rods, cylinders, spheres, and vesicles⁴⁶. In addition, the balance between the hydrophilic and hydrophobic character of the system can be tuned through variation of the relative length of the respective blocks. Stimuli-responsive ELPs have also been achieved by including peptide units that undergo reversible self-assembly driven by well-defined conformational changes when induced by an external stimulus⁷¹⁻⁷³. Because of the versatility of these materials, elastin amphiphilic block copolymers are attractive candidates for medical applications in tissue engineering and therapeutics delivery^{74,75}.

This work aims to develop methods that facilitate the multiple site substitution of amino acid analogues into polypeptide chains. This effort represents a significant advance in biosynthetic technology, promising access to protein-based materials with multiple non-canonical amino acids encoded at defined sites within the polypeptide sequence. Therefore, this system overcomes the limitation of protein biomaterials with twenty-limited natural amino acids, particularly in comparison to organic polymers derived from chemical synthesis and thus suitable for the design of novel biomaterials derived from

sequence-repetitive polypeptides. A universal system for the multi-site selective incorporation of unnatural amino acid analogues into protein by adapting amber suppression strategy is described in chapter 2. We demonstrated the use of the *E. coli* host bacterial system together with the orthogonal pyrrolysyl-tRNA synthetase (PylRS)/tRNA^{pyl}_{CUA} pair to successfully incorporate amino acid analogues into elastin-mimetic polypeptides in response to amber termination codons. In chapter 3, we further extended the system for phenylalanine analogues in which photo-crosslinkers, *para*-benzoyl-L-phenylalanine (Bpa), and *para*-azido-L-phenylalanine (AzF), are selected for incorporation. We optimized the activity of the orthogonal pairs toward the selected analogues for the production of photo-crosslinkable elastin-mimetic polypeptides. In addition, we generated a genetic universal platform for the design of ELP diblock copolymers capable of unnatural amino acids substitution. In chapter 4, we detailed this application in the design of a novel ELP diblock copolymer with photo-crosslinkable activities.

Reference

- [1] A. P. Lothrop, M. P. Torres, S. M. Fuchs, *FEBS Lett.* **2013**, 587, 1247-1257.
- [2] N. Voloshchuk, J. K. Montclare, *Mol. BioSyst.* **2010**, 6, 65-80.
- [3] B. L. Nilsson, R. J. Hondal, M. B. Soellner, R. T. Raines, *J. Am. Chem. Soc.* **2003**, 125, 5268-5269.
- [4] T. Kimmerlin, D. Seebach, *J. Pept. Res.* **2005**, 65, 229-260.
- [5] V. Muralidharan, T. W. Muir, *Nat. Methods* **2006**, 3, 429-438.
- [6] T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel, *Annu. Rev. Biochem.* **2004**, 73, 147-176.
- [7] L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, 35, 225-249.
- [8] J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, D. A. Tirrell, *Curr. Opin. Chem. Biol.* **2010**, 14, 774-780.
- [9] T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel, *Annu. Rev. Biochem.* **2004**, 73, 147-176.
- [10] K. L. Kiick, J. C. M. van Hest, D. A. Tirrell, *Angew. Chem. Int. Ed.* **2000**, 39, 2148-2152.
- [11] K. Kirshenbaum, I. S. Carrico, D. A. Tirrell, *ChemBioChem* **2002**, 3, 235-237.
- [12] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, 79, 413-444.
- [13] L. Davis, J. W. Chin, *Nat. Rev. Mol. Cell Biol.* **2012**.
- [14] M. G. Hoesl, N. Budisa, *Angew. Chem. Int. Ed.* **2011**, 50, 2896-2902.
- [15] T. Hohsaka, Y. Ashizuka, H. Taira, H. Murakami, M. Sisido, *Biochemistry* **2001**, 40, 11060-11064.
- [16] T. Hohsaka, Y. Ashizuka, H. Murakami, M. Sisido, *Nucleic Acids Res.* **2001**, 29, 3646-3651.
- [17] H. Murakami, D. Kourouklis, H. Suga, *Chem. Biol.* **2003**, 10, 1077-1084.
- [18] M. Taki, Y. Tokuda, T. Ohtsuki, M. Sisido, *J. Biosci. Bioeng.* **2006**, 102, 511-517.
- [19] B. M. Eisenhauer, S. M. Hecht, *Biochemistry* **2002**, 41, 11472-11478.

- [20] N. E. Fahmi, L. Dedkova, B. X. Wang, S. Golovine, S. M. Hecht, *J. Am. Chem. Soc.* **2007**, *129*, 3586-3597.
- [21] T. Hohsaka, K. Sato, M. Sisido, K. Takai, S. Yokoyama, *FEBS Lett.* **1994**, *344*, 171-174.
- [22] D. L. Beene, D. A. Dougherty, H. A. Lester, *Curr. Opin. Neurobiol.* **2003**, *13*, 264-270.
- [23] D. A. Dougherty, *Curr. Opin. Chem. Biol.* **2000**, *4*, 645-652.
- [24] E. A. Rodriguez, H. A. Lester, D. A. Dougherty, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8650-8655.
- [25] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498-500.
- [26] T. J. Magliery, J. C. Anderson, P. G. Schultz, *J. Mol. Biol.* **2001**, *307*, 755-769.
- [27] M. Z. Lin, L. Wang, *Physiology* **2008**, *23*, 131-141.
- [28] S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044-1048.
- [29] D. Summerer, S. Chen, N. Wu, A. Deiters, J. W. Chin, P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9785-9789.
- [30] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020-11024.
- [31] S. E. Cellitti, D. H. Jones, L. Lagpacan, X. S. Hao, Q. Zhang, H. Y. Hu, S. M. Brittain, A. Brinker, J. Caldwell, B. Bursulaya, G. Spraggon, A. Brock, Y. Ryu, T. Uno, P. G. Schultz, B. H. Geierstanger, *J. Am. Chem. Soc.* **2008**, *130*, 9268-9281.
- [32] N. Srinivasan, S. Kumar, *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **2012**, *4*, 204-218.
- [33] X. Hu, P. Cebe, A. S. Weiss, F. Omenetto, D. L. Kaplan, *Materials Today* **2012**, *15*, 208-215.
- [34] S. C. Payne, M. Patterson, V. P. Conticello, in *Protein Engineer Handbook, Vol. 2* (Eds.: S. Lutz, U. T. Bornscheuer), Wiley-VCH, **2009**, pp. 915-938.

- [35] A. Takasu, S. Kondo, A. Ito, Y. Furukawa, M. Higuchi, T. Kinoshita, I. Kwon, *Biomacromolecules* **2011**, *12*, 3444-3452.
- [36] J. Raphael, A. Parisi-Amon, S. C. Heilshorn, *J. Mater. Chem.* **2012**, *22*, 19429-19437.
- [37] I. S. Carrico, S. A. Maskarinec, S. C. Heilshorn, M. L. Mock, J. C. Liu, P. J. Nowatzki, C. Franck, G. Ravichandran, D. A. Tirrell, *J. Am. Chem. Soc.* **2007**, *129*, 4874-+.
- [38] O. S. Rabotyagova, P. Cebe, D. L. Kaplan, *Biomacromolecules* **2011**, *12*, 269-289.
- [39] F. G. Omenetto, D. L. Kaplan, *Science* **2010**, *329*, 528-531.
- [40] X. A. Hu, X. L. Wang, J. Rnjak, A. S. Weiss, D. L. Kaplan, *Biomaterials* **2010**, *31*, 8121-8131.
- [41] X. Hu, S. H. Park, E. S. Gil, X. X. Xia, A. S. Weiss, D. L. Kaplan, *Biomaterials* **2011**, *32*, 8979-8989.
- [42] J. Skopinska-Wisniewska, A. Sionkowska, A. Kaminska, A. Kaznica, R. Jachimiak, T. Drewna, *Appl. Surf. Sci.* **2009**, *255*, 8286-8292.
- [43] J. L. Chen, Z. Yin, W. L. A. Shen, X. A. Chen, B. C. Heng, X. A. H. Zou, H. W. Ouyang, *Biomaterials* **2010**, *31*, 9438-9451.
- [44] S. A. Sell, M. J. McClure, K. Garg, P. S. Wolfe, G. L. Bowlin, *Adv. Drug Delivery Rev.* **2009**, *61*, 1007-1019.
- [45] L. Buttafoco, N. G. Kolkman, P. Engbers-Buijtenhuijs, A. A. Poot, P. J. Dijkstra, I. Vermes, J. Feijen, *Biomaterials* **2006**, *27*, 724-734.
- [46] T. A. T. Lee, A. Cooper, R. P. Apkarian, V. P. Conticello, *Adv. Mater. (Weinheim, Ger.)* **2000**, *12*, 1105-1110.
- [47] M. R. Dreher, A. J. Simnick, K. Fischer, R. J. Smith, A. Patel, M. Schmidt, A. Chilkoti, *J. Am. Chem. Soc.* **2008**, *130*, 687-694.
- [48] W. Kim, J. Thevenot, E. Ibarboure, S. Lecommandoux, E. L. Chaikof, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 4257-4260.
- [49] W. Kim, C. Brady, E. L. Chaikof, *Acta Biomater.* **2012**, *8*, 2476-2482.

- [50] W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz, D. A. Tirrell, *Science* **1998**, *281*, 389-392.
- [51] E. R. Wright, V. P. Conticello, *Adv. Drug. Deliv. Rev.* **2002**, *54*, 1057-1073.
- [52] E. R. Wright, R. A. McMillan, A. Cooper, R. P. Apkarian, V. P. Conticello, *Adv. Funct. Mater.* **2002**, *12*, 149-154.
- [53] C. Xu, V. Breedveld, J. Kopecek, *Biomacromolecules* **2005**, *6*, 1739-1749.
- [54] N. Badi, J. F. Lutz, *Chem. Soc. Rev.* **2009**, *38*, 3383-3390.
- [55] J. C. M. van Hest, D. A. Tirrell, *Chem. Commun.* **2001**, 1897-1904.
- [56] K. Nagapudi, W. T. Brinkman, J. E. Leisen, L. Huang, R. A. McMillan, R. P. Apkarian, V. P. Conticello, E. L. Chaikof, *Macromolecules* **2002**, *35*, 1730-1737.
- [57] B. Vrhovski, A. S. Weiss, *Eur. J. Biochem.* **1998**, *258*, 1-18.
- [58] B. A. Cox, B. C. Starcher, D. W. Urry, *J. Biol. Chem.* **1974**, *249*, 997-998.
- [59] D. W. Urry, *J. Protein Chem.* **1988**, *7*, 81-114.
- [60] D. W. Urry, *J. Protein Chem.* **1988**, *7*, 1-34.
- [61] H. M. Kagan, S. N. Gacheru, P. C. Trackman, S. D. Calaman, F. T. Greenaway, in *PQQ and Quinoproteins* (Eds.: J. A. Jongejan, J. A. Duine), Springer Netherlands, Netherlands, **1989**, pp. 317-326.
- [62] K. Reiser, R. J. McCormick, R. B. Rucker, *FASEB J.* **1992**, *6*, 2439-2449.
- [63] D. W. Urry, D. C. Gowda, T. M. Parker, C. H. Luan, M. C. Reid, C. M. Harris, A. Pattanaik, R. D. Harris, *Biopolymers* **1992**, *32*, 1243-1250.
- [64] D. W. Urry, C. H. Luan, T. M. Parker, D. C. Gowda, K. U. Prasad, M. C. Reid, A. Safavy, *J. Am. Chem. Soc.* **1991**, *113*, 4346-4348.
- [65] D. W. Urry, R. D. Harris, M. M. Long, K. U. Prasad, *Int. J. Pept. Protein Res.* **1986**, *28*, 649-660.
- [66] R. S. Rapaka, K. Okamoto, D. W. Urry, *Int. J. Pept. Protein Res.* **1978**, *12*, 81-92.
- [67] D. W. Urry, L. W. Mitchell, T. Ohnishi, *Biochim. Biophys. Acta* **1975**, *393*, 296-306.

- [68] D. E. Meyer, A. Chilkoti, *Biomacromolecules* **2004**, *5*, 846-851.
- [69] X. Wu, R. Sallach, C. A. Haller, J. A. Caves, K. Nagapudi, V. P. Conticello, M. E. Levenston, E. L. Chaikof, *Biomacromolecules* **2005**, *6*, 3037-3044.
- [70] R. E. Sallach, W. Cui, J. Wen, A. Martinez, V. P. Conticello, E. L. Chaikof, *Biomaterials* **2009**, *30*, 409-422.
- [71] L. Martín, E. Castro, A. Ribeiro, M. Alonso, J. C. Rodríguez-Cabello, *Biomacromolecules* **2012**, *13*, 293-298.
- [72] S. R. MacEwan, A. Chilkoti, *Nano Lett.* **2012**, *12*, 3322-3328.
- [73] D. J. Callahan, W. Liu, X. Li, M. R. Dreher, W. Hassouneh, M. Kim, P. Marszalek, A. Chilkoti, *Nano Lett.* **2012**, *12*, 2165-2170.
- [74] W. Kim, J. Xiao, E. L. Chaikof, *Langmuir* **2011**, *27*, 14329-14334.
- [75] V. P. Conticello, H. E. Carpenter Desai, in *Polymer Science: A Comprehensive Reference Vol. 9* (Eds.: K. Matyjaszewski , M. Möller), Elsevier, **2012**, pp. 71-103.

Chapter 2

System development for the multi-site
specific incorporation of non-canonical amino
acids

Introduction

The development of effective biosynthetic methods for the site-specific incorporation of non-canonical amino acids into recombinant polypeptides represents a significant milestone in protein synthetic biology¹. Recently, methods for incorporating non-canonical amino acid residues have advanced substantially for the redesign of natural proteins and *de novo* synthesis of protein analogues with materials applications². Non-native chemical functionality can be introduced at defined sites within native protein sequences in response to unique codons within a genetic template, affording precisely modified protein variants with high efficiency and in good yield. The most powerful approach for site-specifically incorporating unnatural amino acids into proteins expressed in cells is genetic code expansion³. This approach usually involves non-sense suppression of termination codons with a non-natural amino acid analogue. Generally, an aminoacyl-tRNA synthetase (aaRS) and a tRNA are used to specifically insert the unnatural amino acid during mRNA translation, in response to an amber stop codon (UAG) placed at a defined site in a gene. While this methodology has been widely employed, it suffers from a significant limitation in that the target protein yield decreases dramatically as the number of amber-substitution sites increases within the polypeptide sequence due to the multiplicative effect of competition with release factors at each suppression site.

While non-sense suppression is effective for the single insertion of non-canonical amino acids, protein polymers that might have more than one suppression sites within the repetitive sequence are not expected to afford a high-level of substitution, or a high yield of the full-length polypeptide. These protein-based materials have drawn attention for their ability to undergo sequence-directed self-assembly with controlled material

structure and properties⁴⁻⁶. The production of protein-based materials derived from polypeptides of repeating sequences is best accomplished using the techniques of recombinant DNA technology and protein expression^{7,8}. A general scheme for the biological synthesis of repetitive polypeptides used in the construction of protein-based materials involves the 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 polypeptides (Fig. 1). This *in vivo* biosynthesis approach is advantageous over a conventional chemical synthetic strategy where oligopeptides are limited to low degrees of polymerization (< 60 residues) due to the decreasing yield with each successive coupling step. However, protein polymers have a significant disadvantage to conventional organic polymers in that their chemical diversity is generally limited to functional groups that comprise the side-chains of canonical amino acid residues. It would be advantageous to expand the functional group complexity of synthetic protein-based materials beyond that of the canonical amino acids. Recent advances in synthetic biology based on residue-specific incorporation have allowed multiple non-canonical amino acids incorporated at defined positions in the sequence-repetitive polypeptides with a significant yield and degree of substitution². However, this approach has relied primarily on the native *E. coli* biosynthetic machinery and resulted in global substitution of genetically encoded canonical amino acids with structurally similar non-canonical amino acids. The endogenous aaRS displays significantly greater affinity for the native substrate than the non-canonical amino acids, which limits the expression condition in the defined culture media that have been depleted of the canonical amino

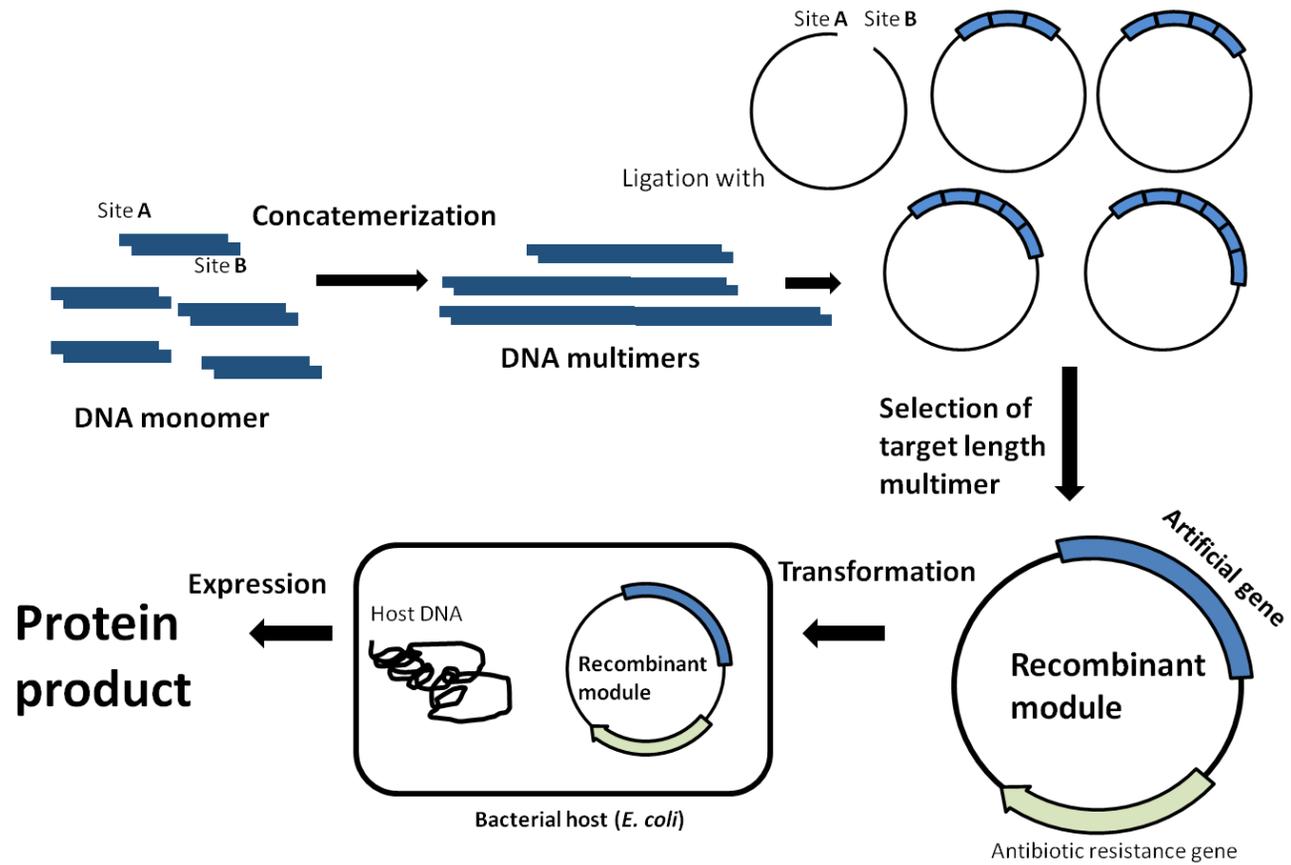


Figure 1. Synthesis of repetitive polypeptides via concatemerizaion of DNA cassettes.

acid. Moreover, this methodology usually requires the use of auxotrophic *E. coli* strains. The isolated yields of the target proteins are variable and the degree of substitution with the non-canonical amino acid is not quantitative⁹.

Recently, several research groups have described approaches for multi-site nonsense suppression in genetically modified *E. coli* strains¹⁰⁻¹³. Although the nonsense codon UAG is the least frequently employed stop codon within the *E. coli* genome, the gene for release factor 1 (RF1), *prfA*, which selectively recognizes the amber codon, is essential for *E. coli* host viability and cannot be eliminated without further modification of the host genome^{10,11}. However, several *E. coli* strains have been described in which mutations have been introduced into the *prfA* gene that attenuate the translation termination activity of RF1 and display enhanced read-through of amber termination codons¹⁴⁻¹⁶. Zhang et al.,¹⁵ reported that a single Arg137Pro mutation within the sequence of *E. coli* RF1 altered its interaction with the ribosome and supported higher levels of amber suppression with genes encoding suppressor tRNA species. Strains with this mutation, *prfA1*, display a temperature-sensitive phenotype, in which growth is impaired at elevated temperature (42 °C). We describe herein the use of one of these strains, *E. coli* MRA30 (MG1655, *recA56 prfA1*),¹⁵ to facilitate the incorporation of multiple non-canonical amino acids into recombinant polypeptides with high degree of substitution. This synthetic approach provides access to the protein polymers that could be modified extensively with non-canonical amino acids at specific positions within the sequence, potentially enhancing the functional repertoire of this novel class of materials.

In addition, pyrrolysine derivatives are selected as targets for the multi-site insertion into the recombinant polypeptides in this study. Pyrrolysine, the 22nd naturally occurring

amino acid encoded by a pyrrolysyl-tRNA synthetase (PylRS)-tRNA pair in archaea species¹⁷, has recently been adapted for encoding various unnatural amino acids in response to an in-frame amber codon in bacteria, mammalian cells, yeast, and *C. elegans*¹⁸⁻²⁰. A unique feature of this pair is its orthogonality in both prokaryotic and eukaryotic organisms that need not be structurally modified to avoid tRNA cross-charging between the heterologous pair and the endogenous system of the host system, making it an attractive facile system for directed evolution in *E. coli* to expand the genetic code with novel amino acid analogues. In addition, the PylRS/tRNA^{pyl}_{CUA} pair comprises a native suppressor in which the CUA anti-codon selectively decodes the amber codon and translationally inserts the variant amino acid with high selectivity²¹⁻²³. Recently, structural modifications were introduced within the substrate-recognition site on the basis of the crystal structure of the PylRS from *Methanosarcina mazei* in order to enhance activity toward pyrrolysine analogues (Fig. 2)²³⁻²⁶. Herein, we constructed a PylRS/tRNA^{pyl}_{CUA} pair from *Methanosarcina barkeri fusaro* as the model orthogonal pair (wt-MbPylRS) for the demonstration of efficient multiple amber suppressions in our system. Structure-based sequence alignments of *M. mazei* PylRS with other PylRSs displayed high similarity of active site residues (Fig. 3)²⁵ and several mutant MbPylRSs were engineered to accommodate pyrrolysine analogues. We constructed the MbPylRS-Y349F mutant that yields higher activity for N^ε-(*tert*-butyloxycarbonyl)-L-lysine (tBoc) and N^ε-allyloxycarbonyl-L-lysine (Alloc) analogues, and the MbPylRS-Y271A/ Y349F mutant with enhanced activity for N^ε-benzyloxycarbonyl-L-lysine (Cbz) analogues (Scheme1). These mutants are suitable for the application of large-scale recombinant

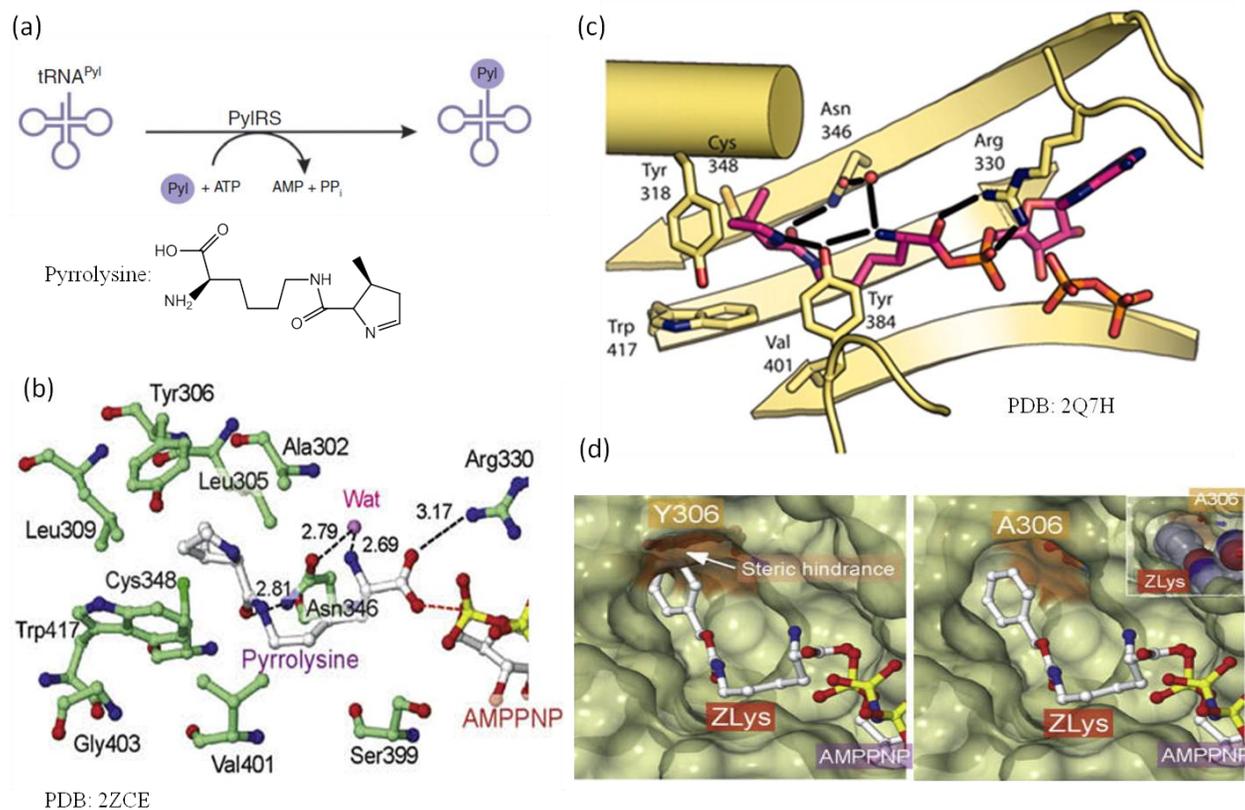


Figure 2. Crystal structures of *M. mazei* PylRS. (a) Schematic representation of PylRS-tRNA^{Pyl} formation and chemical structure of pyrrolysine²³. (b) and (c) Structure of MbPylRS in complex with pyrrolysine^{24,26}. Side chains that interact with pyrrolysine are shown as sticks. The interactions of MbPylRS (yellow) with Pyl-AMP and pyrophosphate (pink) are shown. (d) Docking models of MbPylRS active sites with Cbz-lysine analogue²⁴.

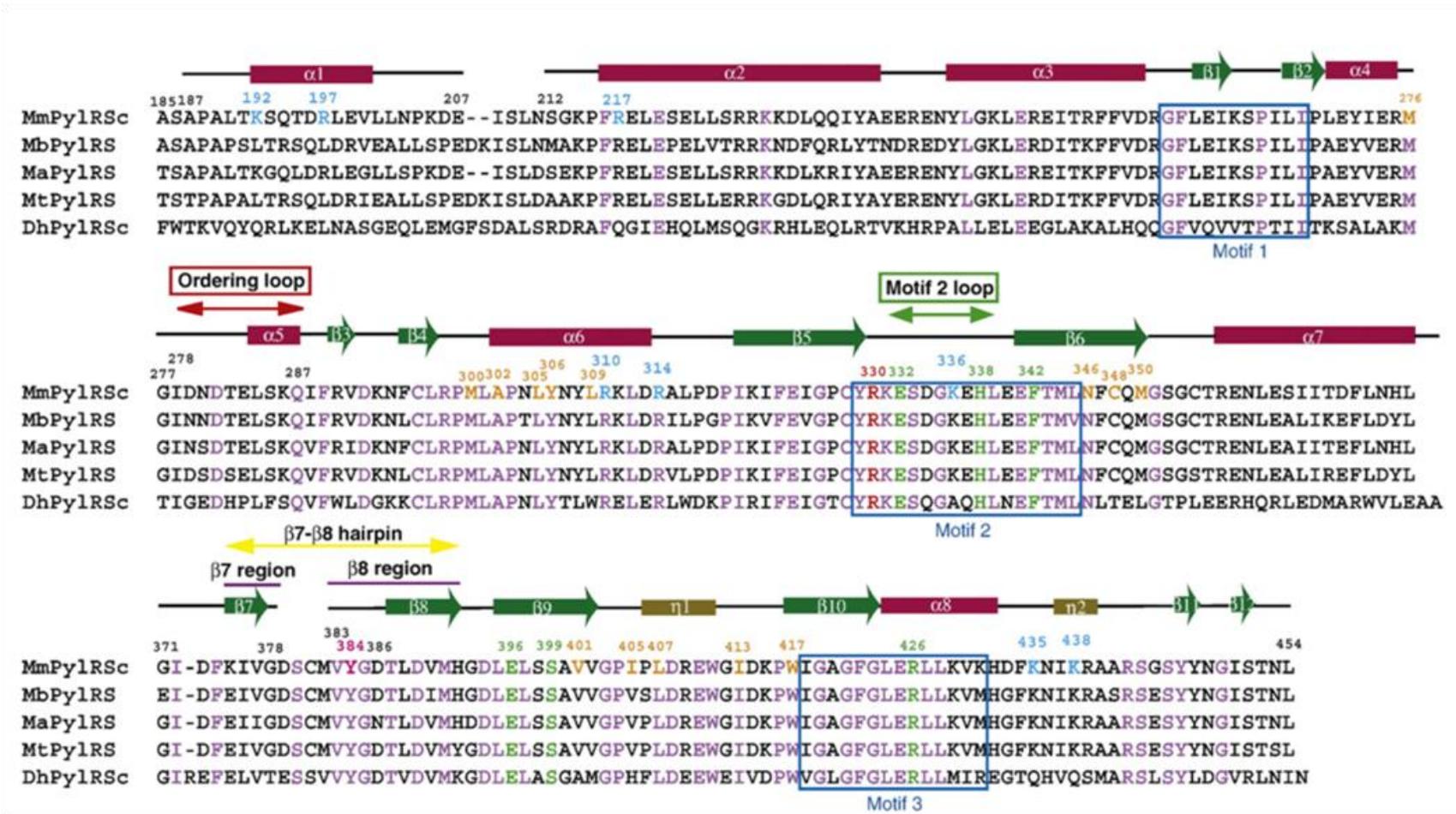
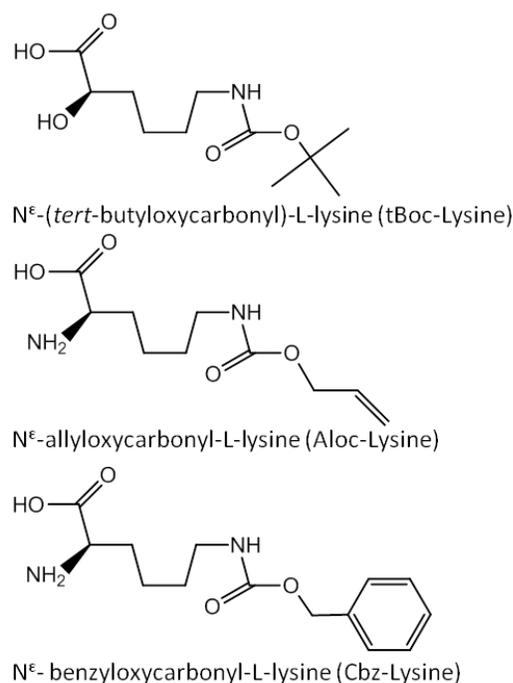


Figure 3. Structure-based sequence alignments of *M. maezi* PylRS with other PylRSs²⁵. Highly conserved residues are shown in purple.

Amino acid residues from the hydrophobic binding pocket for pyrrolysine are shown in orange.

protein expression and the multi-site incorporation of selective pyrrolysine analogues into proteins.



Scheme 1. Pyrrolysine analogues selected as substrates in this study.

The suppression efficiency of our system is assessed by fluorescence-activated cell sorting (FACS). FACS has been successfully used in high-throughput screening to select aaRS variants active with the analogue of interest^{27,28}. Schultz and coworkers demonstrated the installation of amber codons within the T7 RNA polymerase gene fused with the green fluorescent protein (GFP) reporter gene²⁷. Bacterial cells with active synthetase variants expressing GFP were sorted by FACS in a positive and negative screening strategy alternating with antibiotic selection. Tirrell and coworkers also demonstrated a selection method adapting the GFP reporter and FACS to evolve the methionyl-tRNA synthetases for residue-specific incorporation of non-canonical amino

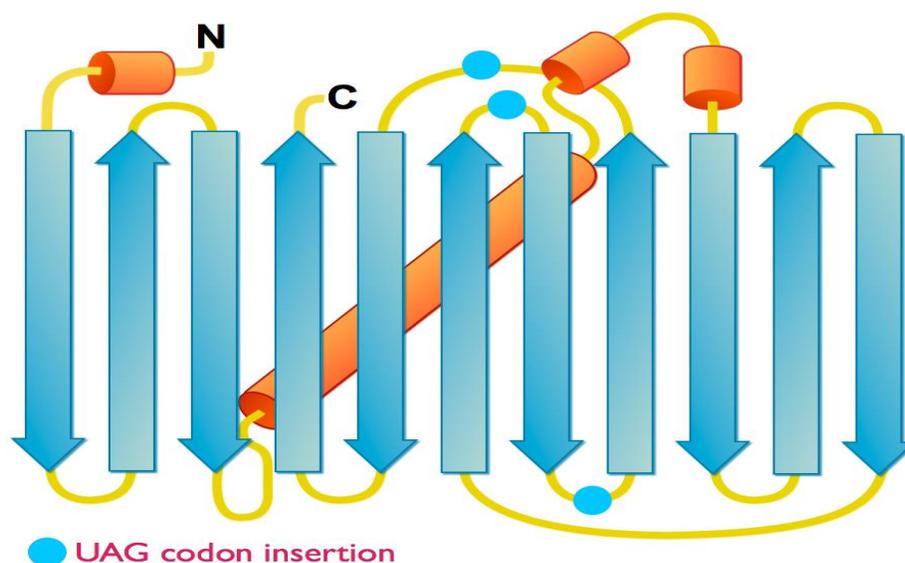


Figure 4. Secondary structure of the sfGFP variant depicting the sites of UAG codon insertion (between 157-158, 172-173, and 194-195). The sites of insertion are selected such that incorporation of the unnatural amino acids would not disrupt proper folding of the sfGFP variant.

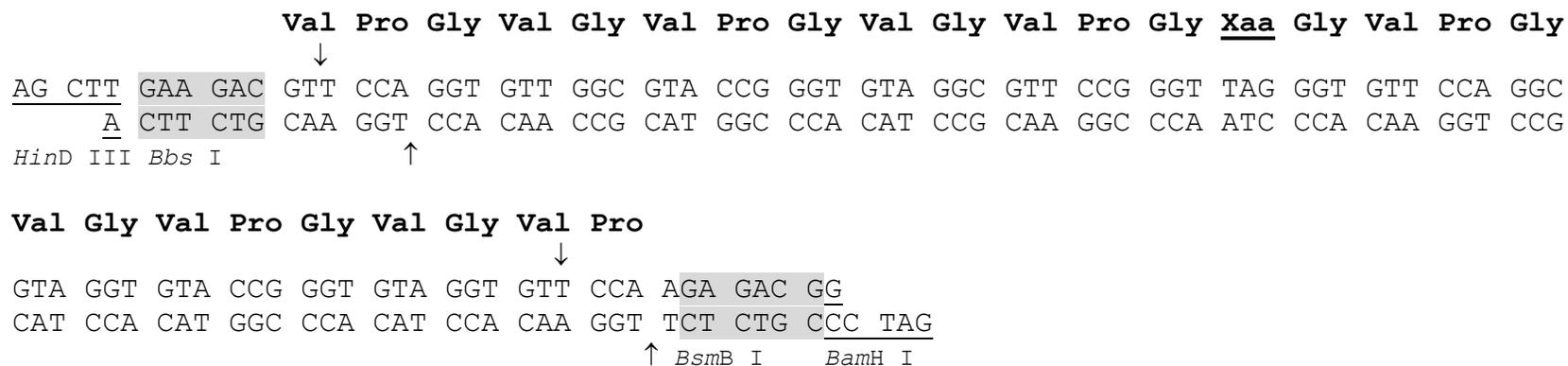
acids in a bacterial host²⁸. Basically, a mutant GFP protein with five additional positions in the primary sequence changed to Met was generated as a reporter for the analogue incorporation. Several rounds of positive and negative screening were conducted using FACS to enrich synthetase variants with a preference for 6,6,6-trifluoronorleucine (Tfn) over the natural Met substrate. Similarly, a fluorescent-based reporter system was constructed in this study to evaluate the suppression efficiency of the MbPylRS pairs in the modified *E. coli* bacterial strains with the presence of non-canonical amino acids. A robustly folded version of GFP, called “superfolder” GFP²⁹ (sfGFP) was generated with three UAG stop codons located in the permissive insertion loop within the GFP coding region (Fig. 4). The loop regions have been identified to accommodate random peptides while allowing the GFP to retain its fluorescence³⁰. Therefore, the improved stability and

folding dynamics of the engineered sfGFP-(UAG)₃ should allow the proper folding of the protein with fluorescence even in the presence of unnatural amino acids. Non-sense suppression of amber codons decoded by the orthogonal system with the incorporation of unnatural amino acids should enable the expression of the full-length sfGFP in cells, which can be detected via its fluorescent emission. In addition, this system that combined the genetic selection of amber codons installed within the sfGFP coding sequence and the fluorescence detection monitored by FACS could be further used in the direction evolution of engineering aaRS variants with altered substrate specificity.

In an effort to design novel protein polymers with substituted amino acid analogues, our lab has demonstrated extensive studies to construct sequence-repetitive polypeptide materials based on the oligopeptide motifs [Val-Pro-Gly-Xaa-Gly] derived from the native elastin³¹⁻³³. The guest residue, which is usually at the fourth position within the pentapeptide unit, has been employed for the global substitution of amino acid analogues such as hydroxyproline³⁴ or *L*-Dopa³⁵ to alter its temperature-dependent phase behavior, mechanical properties, and cross-linked activities. In this chapter, an elastin-mimetic monomer sequence, [(Val-Pro-Gly-Val-Gly)₂(Val-Pro-Gly-Xaa-Gly)(Val-Pro-Gly-Val-Gly)₂] (Scheme 2), was used as the basis for the design of a synthetic DNA cassette in which the introduction of an amber (UAG) codon was employed to specify the position and frequency of occurrence of a non-canonical amino acid (Xaa). A seamless cloning procedure was employed for the head-to-tail self-ligation of the DNA cassette to form a pool of DNA concatemers⁷. Genetic templates based on this repeat sequence can serve as a method to assess suppression efficiency within a defined sequence context, as well as to target the preparation of elastin-like protein (ELP) derivatives with novel

chemical functionality. In addition, this genetic surrogate of ELPs could serve as a universal template for the production of a new set of ELP materials with selective amino acids analogues specified by amber-suppressed orthogonal aaRS pairs.

In this chapter, we demonstrate a simple and efficient method for the introduction of non-canonical amino acids at multiple, structurally defined sites within recombinant polypeptides. The *M. barkeri fusaro* PylRS pairs were employed as the model orthogonal system expressed within the genetically RF1-modified *E. coli* host strain, MRA30. The green fluorescent protein, sfGFP, was constructed as the reporter gene to assess the suppression efficiency in our proposed system. Moreover, we further demonstrate the application of our system with the successful incorporation of unnatural amino analogues into the sequence-repetitive elastin-mimetic polypeptides that have multiple amber codons within the expression cassettes. The synthetic approach described here permits the incorporation of novel chemical groups into protein-based materials at defined positions and expands the versatility of native protein-based materials.



Scheme 2. Oligonucleotide cassettes encoding the **elastin-UAG** monomer 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 II restriction endonucleases.

Experimental Methods

Materials

Chemical reagents were purchased from Fisher Scientific, Inc. (Pittsburgh, PA) or Sigma Chemical Co. (St. Louis, MO) unless otherwise noted. The amino acid derivatives *N*^ε-(*tert*-butyloxycarbonyl)-*L*-lysine (tBoc), *N*^ε-benzyloxycarbonyl-*L*-lysine (Cbz) were purchased Bachem Bioscience, Inc. (King of Prussia, PA), and *N*^ε-allyloxycarbonyl-*L*-lysine (Aloc) was obtained from Sigma Chemical Co. Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, DNA polymerase I large (Klenow) fragment and deoxynucleotide solution mix (dNTPs) were purchased from New England Biolabs, Inc. (Beverly, MA). Plasmid pQE-80L was purchased from Qiagen, Inc. (Valencia, CA). Plasmid pZerO[®]-1 and *Pfx* DNA polymerase were obtained from Invitrogen Corp. (Carlsbad, CA). The pSU81 plasmid was a gift from Professor William McClain at the University of Wisconsin³⁶. Plasmid pHEC2 was synthesized by Dr. Holly Carpenter in the Conticello lab at Emory University³⁴. Single-stranded, PAGE-purified oligonucleotides were obtained from either Sigma-Genosys, Inc. (The Woodlands, TX) or Integrated DNA Technologies (Coralville, IA). Total synthesis of codon-optimized genes encoding wild-type *M. barkeri fusaro* pyrrolysyl-tRNA synthetase and the superfolder GFP variant, sfGFP(UAG)₄, was performed at DNA2.0, Inc. (Menlo Park, CA). The *E. coli* strains, TOP10F' and DH10B were obtained from Invitrogen Corp. (Carlsbad, CA) and the *E. coli* strain MRA30 was kindly provided by permission of Dr. Monica Rydén-Aulin with the assistance of

Jaroslav Belotserkovsky. The plasmids and primers utilized in this chapter are listed in Table 1 and 2 respectively.

General Methods

Basic molecular biology techniques, including polymerase chain reaction, ligation, immunoblotting, and gel electrophoresis, were adapted from a standard molecular cloning manual 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 (0.22 μ m 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. Site-directed mutagenesis was performed using Stratagene's (La Jolla, CA) Quikchange mutagenesis technique from gene-specific oligonucleotide primers. *E. coli* strains were grown in Luria-Bertani (LB) medium with appropriate antibiotics at 37 °C, 200 rpm otherwise noted. Plasmid DNA preparation and purification were performed using the QIAfilter Plasmid Maxi Kit, QIAprep Spin Miniprep Kit (QIAGEN Inc., Valencia, CA), and the DNA Clean & ConcentratorTM (Zymogen, CA). Polymerase chain reaction (PCR) was carried out using a GeneAmp 2400 Thermal Cycler (PerkinElmer Inc. Waltham, MA) and MJ MiniTM Gradient Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules, CA). The PCR products or plasmids were digested with restriction enzymes and separated by 1-4 % agarose TBE gels. The desired DNA fragments were purified with ZymocleanTM Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA). Ligations were carried out in a 10 μ L volume with 0.5 μ L T4DNA Ligase (200 units) and 1X T4DNA Ligase buffer at 16 °C

overnight. Chemically competent cells of TOP10F' or DH10B were transformed with ligation mixtures and recovered in 1 mL SOC rich media at 37 °C for an hour. The recovery mixture was plated onto LB agar media supplemented with appropriate antibiotics and incubated at 37 °C for 12 to 14 hours.

Protein electrophoresis was performed using 10-15% gradient SDS polyacrylamide gels on a PhastSystem from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ) or the Mini-PROTEIN 3 cell electrophoresis system from Bio-Rad Laboratories, Inc. (Hercules, CA). The Perfect Protein Marker (15-150 kDa) (Novagen, Inc., Madison, WI) and Precision Plus Prestained Standards (Bio-Rad Laboratories, Inc., Hercules, CA) were used as protein standards for SDS-PAGE and western blot analysis. Silver staining, used for protein visualization, was carried out using a silver staining kit purchased from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ). Protein expression was detected by western blot using the His-tag AP western Reagents Kit purchased from Novagen[®] (Gibbstown, NJ), which include the secondary antibody, Goat Anti-Mouse IgG AP Conjugate (H + L). Polyhistidine tagged proteins were visualized by chemiluminescent detection using the 1-step NBT/BCIP reagent mixture from Thermo Scientific (Rockford, IL). The sequence of the desired gene was confirmed by automated DNA sequencing analysis using oligonucleotide primers designed to bind upstream and downstream of the target sequence.

Construction of the expression vector

The pQE-80L expression plasmid with the overproducing repressor allele, *lacI_q* is used to ensure tight control of the basal level of transcription prior to induction (Fig. 5). The internal *Nco* I endonuclease restriction site was removed by Klenow-fill method to

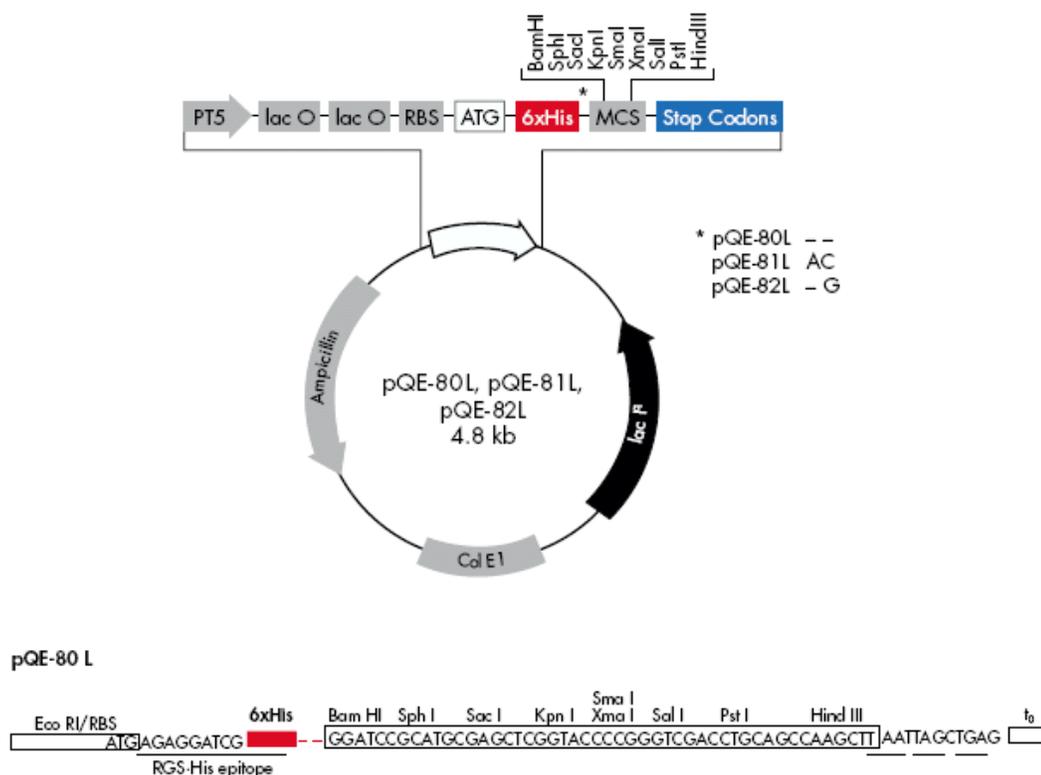


Figure 5. Plasmid map of the pQE-80L vector with multiple cloning site (MCS).

generate pIL1. In brief, the plasmid pQE-80L was digested by *Nco* I and purified by Zymoclean™ Gel DNA Recovery Kit. The overhang ends were filled in with the DNA polymerase I, large (Klenow) fragment in 33μM dNTP solution at 25 °C for 2 h. The linear blunt-end plasmid was ligated by Quick Ligation™ Kit (New England Biolabs, Inc). The plasmid DNA was purified and transformed into TOP10F' cells and the recovery mixture was spread on LB agar supplemented with ampicillin (100 μg/mL). Four single colonies were selected to inoculate LB media supplemented with ampicillin and grown overnight at 37 °C. The recombinant plasmids were confirmed by DNA sequencing and *Nco* I digestion. An adaptor sequence with extra restriction sites and a

deca-histidine tag at the C-terminus was designed for cloning of the following genes in the pQE-80L expression plasmid (Scheme 3). The single-stranded oligonucleotides corresponding to the sense and anti-sense strands of the DNA adaptor, **pQE-Adaptor**, were chemically synthesized. Annealing was carried out by dissolving the primers in sterile ddH₂O to a final concentration of 0.5 µg/µL. Twenty microliter aliquots of each of the two primers were mixed together with 4 µL 5 M NaCl, 4 µL 1 M MgCl₂, and 152 µL sterile ddH₂O. By gradually decreasing the temperature of the reactions from 99 °C to 30 °C (decreased by 1 °C every 3 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 in 33µM dNTP solution at 25 °C for 2 h. The DNA duplex fragment was visualized by DNA agarose gel electrophoresis (4 % NuSieve agarose) and purified using the QIAquick PCR purification Kit. The adaptor and commercially available pQE-80L were double-digested with *EcoR* I and *Hind* III endonuclease enzymes, separated via gel electrophoresis, and purified by Zymoclean™ Gel DNA Recovery Kit individually. The digested adaptor DNA and pQE-80L were ligated together with a 3:1 molar ratio. The ligation reaction was transformed into TOP10F' cells and the recovery mixture was spread on LB agar supplemented with ampicillin (100 µg/mL). The sequence of the recombinant plasmid clones was confirmed by DNA sequencing and designated as pIL2.

For subsequent cloning of the elastin-UAG concatemers, it was necessary to remove the *Bsa* I restriction endonuclease recognition site from the modified expression plasmid. Site-directed mutagenesis was utilized to introduce a single mutation into the

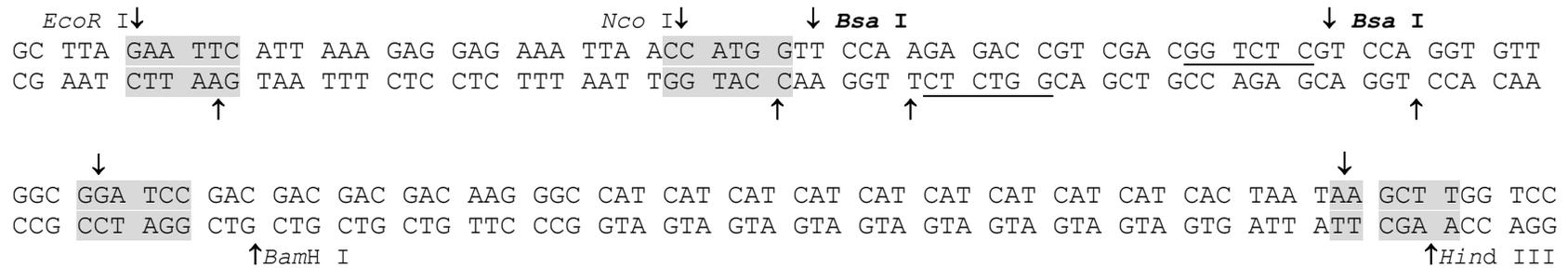
pQE-Adaptor-F

5' -GCTTAGAATTCATTAAAGAGGAGAAATTAACCATGGTTCCAAGAGACCGTCGACGGTCTCGTCCAGGTGTTGGCGGATCCGACGACG-3'

pQE- Adaptor-R

5' -GGACCAAGCTTATTAGTGATGATGATGATGATGATGATGATGATGGCCCTTGTCGTCGTCGTCGGATCCGCCAACACCTGGACGAGA-3'

Adaptor sequence after annealing and Klenow fill-in



Scheme 3. Sequence of pQE-Adaptor forward and reverse primer before and after annealing is shown. The recognition sites for the relevant restriction endonucleases, which are employed for cloning of the target proteins expression plasmid, are highlighted. Arrows indicate the cleavage positions on the sense and anti-sense strand for the respective endonucleases.

sequence of the plasmid by mutagenesis primers, pQE80LBsaI-F and pQE80LBsaI-R, and produce the plasmid pIL5 (Fig. 6).

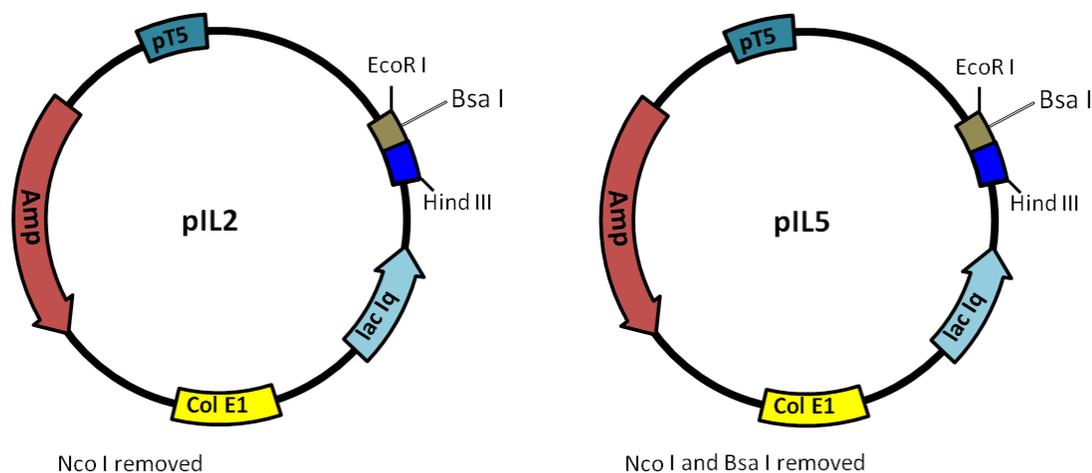


Figure 6. Modified pQE-80L plasmids with pQE-adaptor sequence. The pIL2 plasmid was generated from pIL1 by cloning into the adaptor sequence. The pIL5 was generated by site-directed mutagenesis of pIL2 to remove a *Bsa* I restriction site to enable cloning into the expression plasmid.

Construction of sfGFP expression plasmids

A gene encoding wild-type superfolder GFP was optimized for expression in *E. coli* and synthesized by DNA2.0 (Menlo Park, CA) to produce pJ201-sfGFPPCC. The gene was isolated through *EcoR* I/ *Bam*H I digestion and cloned into digested pIL2 to produce pIL7. A synthetic gene encoding a superfolder green fluorescent protein variant, sfGFP-(UAG)₄, with a single K3stop mutation and three amber codons inserted at sites between amino acid residues 157-158, 172-173, and 194-195, was optimized for expression in *E. coli* and synthesized by DNA2.0 (Menlo Park, CA) to produce pJ201-

sfGFP4TAG. The gene was isolated by *Bam*H I/ *Eco*R I digestion and cloned into the compatible sites in the modified pQE-80L expression plasmid pIL2 as pIL77. The K3stop silent mutation was removed by site-directed mutagenesis utilizing sfGFP3TAG-F/ R primers. The sequence of the expression cassette was verified after isolation of the plasmid and designated as pIL80. In this way, the sfGFP expression cassettes were cloned upstream of an in-frame deca-histidine tag at the C-terminus of the full-length reading frame which facilitates the visualization of the expressed protein by western blot and the purification by metal affinity chromatography (Fig. 7).

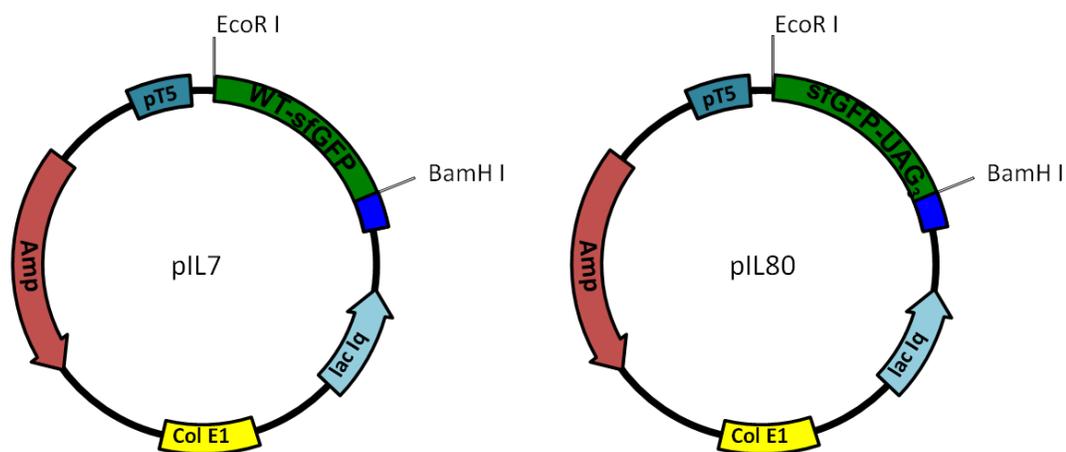


Figure 7. Plasmids pIL7 and pIL80 encode wild-type sfGFP and sfGFP-(UAG)₃.

Construction of the elastin-mimetic peptides, Elastin-UAG genes

The DNA oligonucleotide primers encoding the sense and anti-sense monomer sequence were chemically synthesized and annealed to produce duplex monomer DNA. The synthetic DNA monomer encoding the sequence [Val-Pro-Gly-Val-Gly]₂[Val-Pro-Gly-(amber)-Gly]-[Val-Pro-Gly-Val-Gly]₂ (**elastin-UAG**) was prepared by annealing the forward and reverse oligonucleotide strands, Elastin-UAG-F and Elastin-UAG-R.

Annealing was carried out by dissolving the primers in sterile ddH₂O to a final concentration of 0.5 µg/µL. Twenty microliter aliquots of each of the two primers were mixed together with 4 µL 5 M NaCl, 4 µL 1 M MgCl₂, and 152 µL sterile ddH₂O. By gradually decreasing the temperature of the reactions from 99 °C to 30 °C (decreased by 1 °C every 3 min), the DNA strands were annealed together. The duplex DNA monomer was phosphorylated at the 5' and 3' positions by T4 polynucleotide kinase. The double-stranded DNA fragment was separated and purified by DNA agarose gel electrophoresis (4 % NuSieve agarose). A ligation between the monomer DNA and *Bam*H I/ *Hind* III-digested pZErO[®]-1 (Fig. 8) plasmid was performed and transformed into *E.coli* stain Top10F' chemical competent cells, and selected on low salt LB agar supplemented with 1 mM IPTG and Zeocin[®] (50 µg/mL). Recombinant plasmids were screened via colony PCR using SP6 and T7 primers to identify the elastin-UAG monomer gene within the plasmid. The sequence of the elastin-UAG monomer was confirmed by automated DNA sequencing analysis using M13 reverse and M13 (-20) forward primers. The plasmid, pIL41, is comprised of the pZErO[®]-1 plasmid with the elastin-UAG monomer insert flanked by *Bbs* I and *Bsm*B I restriction enzyme sites.

The elastin-UAG multimer, which encodes a repetitive elastin-mimetic polymer, was constructed using a previous reported DNA concatemerization strategy⁷. DNA cassette concatemerization was performed by first preparing a large amount of elastin-UAG monomer plasmid DNA using the QIAfilter Plasmid Maxi Kit. The plasmid DNA was digested with *Bbs* I for 16 h at 37 °C. The DNA was purified by extraction with phenol/chloroform/isoamyl alcohol (25:24:1) to remove proteins. Ethanol precipitation was used to purify the DNA and reduce the salt concentration. The entire isolated DNA

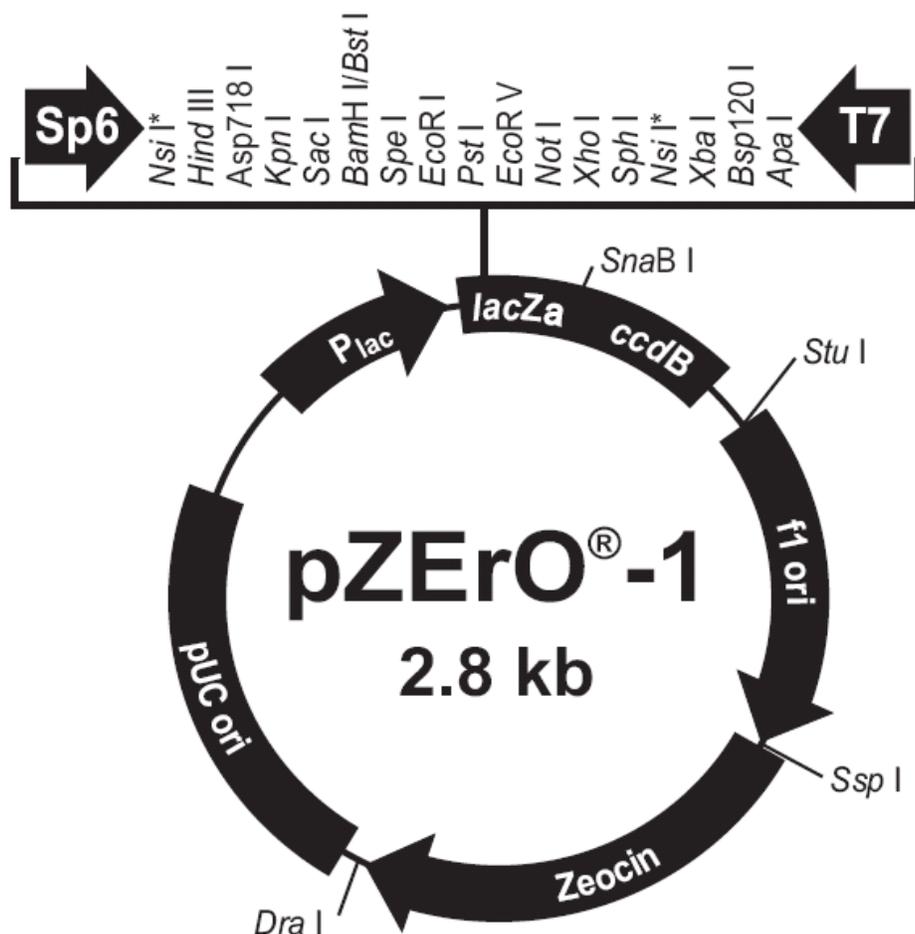


Figure 8. Plasmid map of the pZErO⁻¹ vector, indicating restriction endonuclease cleavage sites within the multiple cloning site. The **elastin-UAG** monomer DNA sequence was cloned into the plasmid using the *BamH I* / *Hind III* sites.

sample was digested with *BsmB I* at 55 °C for 16 h and loaded onto a preparative 4% NuSieve agarose gel in order to separate the digested pZErO⁻¹ plasmid from the elastin-UAG monomer. The monomeric DNA cassette corresponding to the 75 bp with complementary and non-palindromic cohesive ends was purified using the Amicon Ultra free Maximum Recovery Kit (Millipore, Burlington, MA). Self-ligation of the DNA monomer in a head-to-tail fashion was conducted by incubating with T4 DNA ligase and

with ATP at 16 °C for 16 h. The library of DNA concatemers was fractionated via agarose gel electrophoresis (1% agarose) (Fig. 9). The region of the gel corresponding to the desired multimer size range (~1000 to 2000 bp) was excised and purified from the gel using the Zymoclean™ Gel DNA Recovery Kit.

The purified DNA concatemers were ligated back into the *Bbs* I-digested pIL41 parent plasmid containing one copy of the **elastin-UAG** monomer in pZErO®-1. The *Bbs* I-digested pIL41 acceptor plasmid was first dephosphorylated at the 5'-ends with antartic phosphatase at 37 °C for 30 min followed by heat inactivation at 65 °C for 5 min. The ligation mixture was used to transform electrocompetent cells of *E. coli* strain TOP10F' and the transformants were selected on low salt LB agar supplemented with 1 mM IPTG and Zeocin™ (50 µg/mL). The recombinant plasmids were screened via colony PCR with SP6 and T7 primers, and analyzed via agarose gel electrophoresis for the presence of multimer inserts. From the pool of transformants, two recombinant plasmids were isolated with concatemer inserts of approximately 1200 and 2000 bp. The recombinant plasmids containing 12 and 22 repeats of the elastin-UAG monomer were designated as pIL42.5 and pIL42.11, respectively. The sequence of the recombinant plasmids was confirmed by automated DNA sequencing analysis using the M13 reverse and M13 (-20) forward primers. Next, an acceptor plasmid suitable for protein expression was required to conduct high-level expression of elastin-UAG multimers in a bacterial host.

Construction of the Elastin-UAG expression plasmids

The pZErO®-1 cloning plasmids, pIL42.5 and pIL42.11, containing the corresponding elastin-UAG multimer gene were sequentially digested with *Bbs* I and *BsmB* I to liberate the multimer cassettes. The pIL5 plasmid, containing the adaptor

sequence, was digested with *Bsa* I to generate the complementary sticky ends necessary for the seamless cloning of the elastin-UAG multimer genes into the adaptor of this modified pQE-80L expression plasmid. The *Bsa* I-digested pIL5 was dephosphorylated with antartic phosphate to reduce the false positive self-ligations. Ligation reactions were incubated at 16 °C for 12 h to covalently link the digested pIL5 and the digested elastin-UAG multimers (12 and 22 repeats). The ligation mixtures were 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 the transformants and screened via double digestion with *Nco* I/ *Hind* III for the presence of the elastin-UAG concatemer inserts. The sequences of the elastin-UAG gene multimers was confirmed by automated DNA sequencing analysis using PQE-F/ R primers. The correct plasmids, pIL43.5 and pIL43.11, included the pQE-80L plasmid with **Elastin-UAG** gene insert approximately 1200 (12 repeats) and 2000 bp (22 repeats) in length, respectively (Fig. 10).

Construction of the orthogonal aminoacyl-tRNA synthetase/tRNA pairs

***M. barkeri fusaro* pyrrolysyl-tRNA synthetase/tRNA (MbPylRS/tRNA^{pyl}_{CUA}) pair**

The pHEC2 plasmid synthesized by Dr. Holly Carpenter³⁴ previously was used for the expression of MbPylRS/tRNA^{pyl}_{CUA} pairs. The pHEC2 plasmid was derived from the pPROTet/Lar plasmid containing the transcriptional/translational control elements³⁷ for the expression of MbPylRS and multiple cloning sites for the cloning of tRNA gene and synthetase gene (Fig. 11).

A duplex DNA cassette encoding MbtRNA_{CUA} gene was constructed by annealing the synthetic oligonucleotides primers, pylTtRNA-F and pylTtRNA-R. The

primers were designed to incorporate *EcoR* I and *Pst* I sticky ends, 4 nucleotide base overhangs, for direct cloning of the tRNA gene into the acceptor plasmid, pSU81. The DNA cassette was phosphorylated by T4 polynucleotide kinase at 37 °C for 1 hour and purified by Zymogen DNA Clean & Concentrator™ kit. The phosphorylated MbtRNA_{CUA} gene was then cloned into compatible *EcoR* I and *Pst* I restriction sites within plasmid pSU81, transformed into the chemical competent cells of *E. coli* strain Top10F', and spread onto LB agar supplemented with chloramphenicol (34 µg/mL). The pSU81 plasmid, containing the MbtRNA_{CUA} gene flanked by the *lpp* tRNA promoter and *rrnC* tRNA terminator sequences was sequenced by using pSU81-F/ R primers and assigned as pIL38. The pIL38 plasmid was double digested with *Nhe* I and *Pvu* I to afford the MbtRNA_{CUA} gene as a duplex DNA cassette under the transcriptional control of the *lpp* promoter and *rrnC* terminator sequences. The DNA encoding the tRNA expression cassette was cloned into the compatible sites, *Nhe* I and *Pvu* I, within plasmid pHEC2 to generate plasmid pIL39 and confirmed by automated DNA sequencing analysis using HtRNA-F/ R primers.

The synthetic gene encoding the wild-type MbPylRS was optimized for the expression in *E. coli* and synthesized by DNA2.0 (Menlo Park, CA) to produce pJ201-MbPylRS. The MbPylRS gene was released by double digestion with *Kpn* I/ *Xba* I and cloned into the acceptor plasmid pIL39 that already contained the MbtRNA_{CUA} gene expression cassette. The plasmid was assigned as pIL40 and sequenced by 5-Pro/ 3-Pro primers. The wild-type MbPylRS gene in the final expression plasmid, pIL40, is under the control of the constitutive $P_L tet$ promoter (Fig. 12).

***M. barkeri fusaro* pyrrolysyl-tRNA synthetase/tRNA (MbPylRS/tRNA^{Pyl}_{CUA}) mutants**

Several mutants based on crystal structures of *M. mazi* PylRS (PDB: 2Q7E, 2Q7G, 2Q7H) were reported to accommodate various pyrrolysine analogues. Single (Y271A, Y349F) and double (Y271A/ Y349F) mutations were introduced into the wild-type MbPylRS gene using gene-specific primers. The PCR products were digested with *Dpn* I at 37 °C for overnight and purified by DNA Clean & ConcentratorTM (Zymogen, CA). The DNA was transformed directly into *E. coli* strain Top10F', recovered, and plated onto LB agar supplemented with chloramphenicol (34 µg/mL). Recombinant plasmids were selected and sequenced by 5-Pro/ 3-Pro primers to confirm the location of mutation. The pIL50, pIL51, and pIL58 plasmids contained single (Y349F, Y271A) and double (Y271A/Y349F) mutations respectively (Fig. 13).

Protein expression and purification

Small scale expression of sfGFPs in DH10B strain

The expression vector pIL80 (sfGFP-(UAG)₃ gene) was co-transformed with pIL40 (wild-type MbPylRS pair) or pHEC2 (lacking the synthetase/tRNA pair), into the DH10B strain and plated on LB agar media supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL). The pIL80 plasmid DNA was also co-transformed with other mutant MbPylRS pairs into the DH10B strain and plated on LB agar supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL).

Single colonies of the expression strains were inoculated into sterile LB broth supplemented with the appropriate antibiotics (100 µg/mL ampicillin and 34 µg/mL chloramphenicol) as required for plasmid maintenance. The overnight culture was diluted

50-fold the next morning into fresh media, and cells were allowed to grow at 37 °C until reached log phase (OD_{600} reached approximately 0.6-0.8). A final concentration of 1mM IPTG and amino acid analogues were added to induce the recombinant protein expression. The cultures were incubated at 37 °C for 12 hr and analyzed by SDS-PAGE and western blot for evidence of the protein synthesis. Briefly, samples were prepared by mixing aliquots of the sample culture with 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 μ L 1 M DTT, and sterile ddH₂O to a final volume of 25 μ L and boiling mixture at 100 °C for 5 min.

Expression and purification of sfGFPs in MRA30 strain

The expression vector pIL80, encoding sfGFP-(UAG)₃, was first transformed into electro-competent cells of *E. coli* MRA30 strain and plated on LB agar supplemented with ampicillin (100 μ g/mL). Single colony of the MRA30 strain bearing pIL80 was inoculated and made as electro-competent cells for further transformation. Electroporation was used to transform the DNA as this method affords the highest transformation efficiency compared to chemical transformation. The pIL40 (wild-type MbPylRS pair) or pHEC2 (lacking the synthetase/tRNA pair), or other MbPylRS mutant pair was transformed into the MRA30 strain containing pIL80 plasmid already. The recovery mixture was plated onto LB agar supplemented with ampicillin (100 μ g/mL) and chloramphenicol (34 μ g/mL) and incubated at 30 °C for 12 to 14 hr.

Single colonies of the expression strains were inoculated into sterile LB broth supplemented with the appropriate antibiotics as required for plasmid maintenance. One liter of LB supplemented with the appropriate antibiotics was inoculated with 20 mL of

the overnight culture and incubated at 30 °C until the OD₆₀₀ reached between 0.6 and 0.8. Pyrrolysine analogues and IPTG were added to a final concentration of 1 mM to induce the expression of the sfGFP derivatives. After 12 h, the cells were harvested by centrifugation at 4000 g and 4° C for 20 min. The cell pellet was re-suspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 8.0) and stored at –80°C. The frozen cells were lysed by three freeze/ thaw cycles. Lysozyme (1 mg/mL), EDTA-free protease inhibitor cocktail, benzonase (25 units/mL), and MgCl₂ (1 mM) were added to the lysate and the mixture was incubated shaking at 4 °C for 12 h. The cell lysate was centrifuged at 14,000 g for 30 min at 4 °C. SfGFPs were purified by metal-affinity chromatography using the cobalt-charged TALON[®] resin from Clontech Laboratories (Mountainview, CA). The supernatant from the sfGFP expression was loaded directly onto a column containing cobalt-charged TALON[®] resin (2 mL) and washed with lysis buffer (20 mL) containing 20 mM imidazole. The target protein was eluted with elution buffer (20 mL, 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluate was concentrated by ultrafiltration using Amicon filters (NMWL 10 kDa) to a final volume of 1 mL and desalted by desalting columns (Thermo Scientific, Inc.). Protein concentration was determined by the Bradford assay.

Small scale expression of Elastin-UAG in DH10B strain

Elastin-(UAG)₁₂ (pIL43.5) or elastin-(UAG)₂₂ (pIL43.11) plasmid was co-transformed with pIL40 (wild-type MbPylRS pair) or pHEC2 (lacking the synthetase/tRNA pair), or other MbPylRS mutant pairs into the DH10B strain and plated on LB agar supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL). Single colonies of the expression strains were inoculated into sterile LB broth

supplemented with the appropriate antibiotics as required for plasmid maintenance. The overnight culture was diluted 50-fold into fresh media the next morning. The cell cultures were allowed to grow at 37 °C until the cells reached log phase growth (OD_{600} 0.6 to 0.8). A final concentration of 1mM IPTG and amino acid analogues were added to induce the recombinant protein expression. The expression cultures were incubated at 37 °C for 12 h and the protein expression was monitored by SDS-PAGE and western blotting.

Expression and purification of Elastin-UAG in MRA30 strain

Electro-competent cells of *E. coli* strain MRA30 were prepared and transformed with the plasmids pIL43.5 or pIL43.11, encoding the target proteins, and selected on LB agar supplemented with ampicillin (100 µg/mL) via incubation at 30 °C for 12-14 h. Fresh electro-competent cells were prepared from single colonies for further transformation with MbPylRS pairs to generate the expression strains employed for these studies. The transformation mixture was plated onto LB agar supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) at 30 °C for 12-14 h.

Single colonies of each expression strain were inoculated into sterile LB broth supplemented with the appropriate antibiotics as required for plasmid maintenance. One liter of LB supplemented with the appropriate antibiotics was inoculated with 20 mL of overnight culture and incubated at 30 °C until the OD_{600} reached between 0.6 and 0.8. Pyrrolysine analogues and IPTG were added to a final concentration of 1 mM to induce the expression of the elastin-mimetic proteins. After 12 h, the cells were harvested by centrifugation at 4 °C (4000 g) for 20 min. The cell pellet was re-suspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and stored at -80°C.

The frozen cells were lysed by three freeze/ thaw cycles. Lysozyme (1 mg/mL), EDTA-free protease inhibitor cocktail, benzonase (25 units/mL), and MgCl₂ (1mM) were added to the lysate and the mixture was incubated shaking at 4 °C for 12 h. The cell lysate was centrifuged at 14,000 g for 30 min at 4 °C. The elastin-mimetic polymers were purified through repetitively cycling through the inverse temperature transition between 30 °C and 4 °C. A final concentration of 2 M NaCl was added into the initial supernatant from the elastin protein expression cultures to induce phase separation above the lower critical solution temperature (LCST). The mixture was incubated at 30 °C for 30 min to precipitate the target protein. At this temperature, the elastins derivatives became insoluble and formed large micron-sized aggregates. The suspension was centrifuged at 9500 g and 25 °C for 20 min. The pellet containing the precipitated elastin derivative was re-suspended in cold (4 °C) lysis buffer, incubated on ice for 30 min and centrifuged at 20000 g and 4 °C for 20 min. The supernatant was analyzed by 10-15 % gradient SDS-PAGE to assess the purity of the elastin derivatives. This thermal cycling process was repeated as previously described until elastin derivatives were purified from the endogenous host proteins, usually three to five repetitions. The purity of elastins was assessed via SDS-PAGE and the pure elastins were dialyzed against distilled deionized water (5x 4 L). The dialysate was lyophilized to produce colorless solids and the protein yields were measured in terms of dry weight per unit volume of culture.

Flow cytometry

Aliquots (1 mL) of *E. coli* cells from expression cultures were grown until the OD₆₀₀ reached approximately 1.0. The cultures were centrifuged at 4000g and 4 °C for 10 min and re-suspended in 1 mL of phosphate-buffered saline pH 7.4. Flow cytometry was

performed using a LSRII flow cytometer (Beckton Dickinson) equipped with a 100 mW solid-state laser emitting at 488 nm for the excitation of sfGFP, a 505 nm LP dichroic mirror and a 530/30 bandpass filter. Forward scatter (FSC), sideward scatter (SSC) and green fluorescence 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 with different non-canonical amino acids. Data was analyzed using FlowJo software (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 a final concentration of 10 mM and the mixture was incubated at 100 °C for 30 min to denature the protein. After the reaction mixture was cooled to 37 °C, thermolysin was added to the denatured protein solution to a concentration of 1: 50 (w/w) ratio with respect to protein and the reaction mixture was incubated at 37 °C for 12 h. The products from the proteolysis reaction were passed through a PepCleanTM C18 spin column (Thermo Scientific, Inc.) to remove salts and thermolysin. Briefly, samples were adjusted to a final concentration of 0.5% trifluoroacetic acid (TFA) in 5 % acetonitrile (ACN). Each spin column was activated by adding 50% ACN and centrifuged at 1500 g for 1 min (2x 200 µL). The spin column was then equilibration with 0.5% TFA in 5 % ACN (2x 200 µL). Digested samples were loaded onto the spin column and centrifuged at 1500 g for 1 min. The loading step was repeated at least twice to ensure complete binding. The column was washed with 0.5%

TFA in 5 % ACN (3x 200 μ L). The peptide fragments were eluted by 70 % ACN (2x 20 μ L) and the solution was employed directly for mass spectrometric analysis.

Mass spectrometry

Electrospray spectra of sfGFP derivatives were acquired by LTQ-FT mass spectrometer (ThermoElectron, San Jose, CA) in the positive ion mode (200-2000 m/z) using a spray voltage of 4 kV, a capillary voltage of 41 V, a sheath gas flow rate of 20 arbitrary units. The tube lens voltage was 215V, and the AGC setting was 5e+05. The Xtract program in Xcalibur (ThermoElectron, San Jose, CA) was used for spectral deconvolution. MALDI-TOF experiments were performed on an Applied Biosystems[®] Voyager[™] System 428 mass spectrometer (Life Technologies Corporation; Carlsbad, CA) in the positive ion linear mode. Ferulic acid was used as matrix at a concentration of 10 mg/mL in a mixture of 75 % ACN and 0.2% FA in deionized water. The elastin-mimetic protein solution (1 mg/mL in distilled water) was mixed with the matrix solution in a ratio of 1:10 and vortexed gently for a few seconds. Two-microliters of the mixture were spotted on a stainless steel, 100-position flat sample plate and dried under air. The Sequazyme[™] bovine serum albumin (BSA) (Applied Biosystems Inc.; Carlsbad, CA) was used as a standard for external calibration with the sinapinic acid matrix solution. The sinapinic acid (3-(4-hydroxy-3, 5-dimethoxyphenyl)prop-2-enoic acid) was used at a concentration of 20 mg/mL in a mixture of 50 % ACN and 0.1% FA in deionized water.

Tandem MS of selected thermolysin-digested peptides were carried out using ESI mass spectrometry. Electrospray measurements were acquired by LTQ-FT mass spectrometer (ThermoElectron, San Jose, CA) in positive ion mode using a spray voltage of 4 kV, a capillary voltage of 44 V, a sheath gas flow rate of 20 arbitrary units, a CID

collision energy 25-35eV. The tube lens voltage was 120 V, and the AGC setting was $1e+06$. Solutions of each peptide (1×10^{-5} M, methanol) were infused with a rate of 5 $\mu\text{L}/\text{min}$. Ultra pure helium was the collision gas in MS^n experiments carried out in the ion trap. Thirty to fifty scans were recorded and averaged for accurate mass measurements.

Temperature-dependent turbidity

The thermal transitions of the elastin-mimetic peptides with pyrrolysine analogues were estimated from temperature turbidity profiles. Measurements were performed using a JASCO V-530 UV/Visible spectrophotometer equipped with a programmable Peltier cell and a water-controlled JASCO ETC-505T temperature controller (Jasco, Inc.; Easton, MD). Dilute solutions of the elastin-mimetic peptide samples (1.0 to 3.0 mg/mL) were prepared at 4 °C and analyzed using a quartz cuvette with a 10 mm pathlength (Hellma Analytics; Müllheim, Germany). The absorbance of the polymer solutions was monitored over a temperature range from 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 by 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.

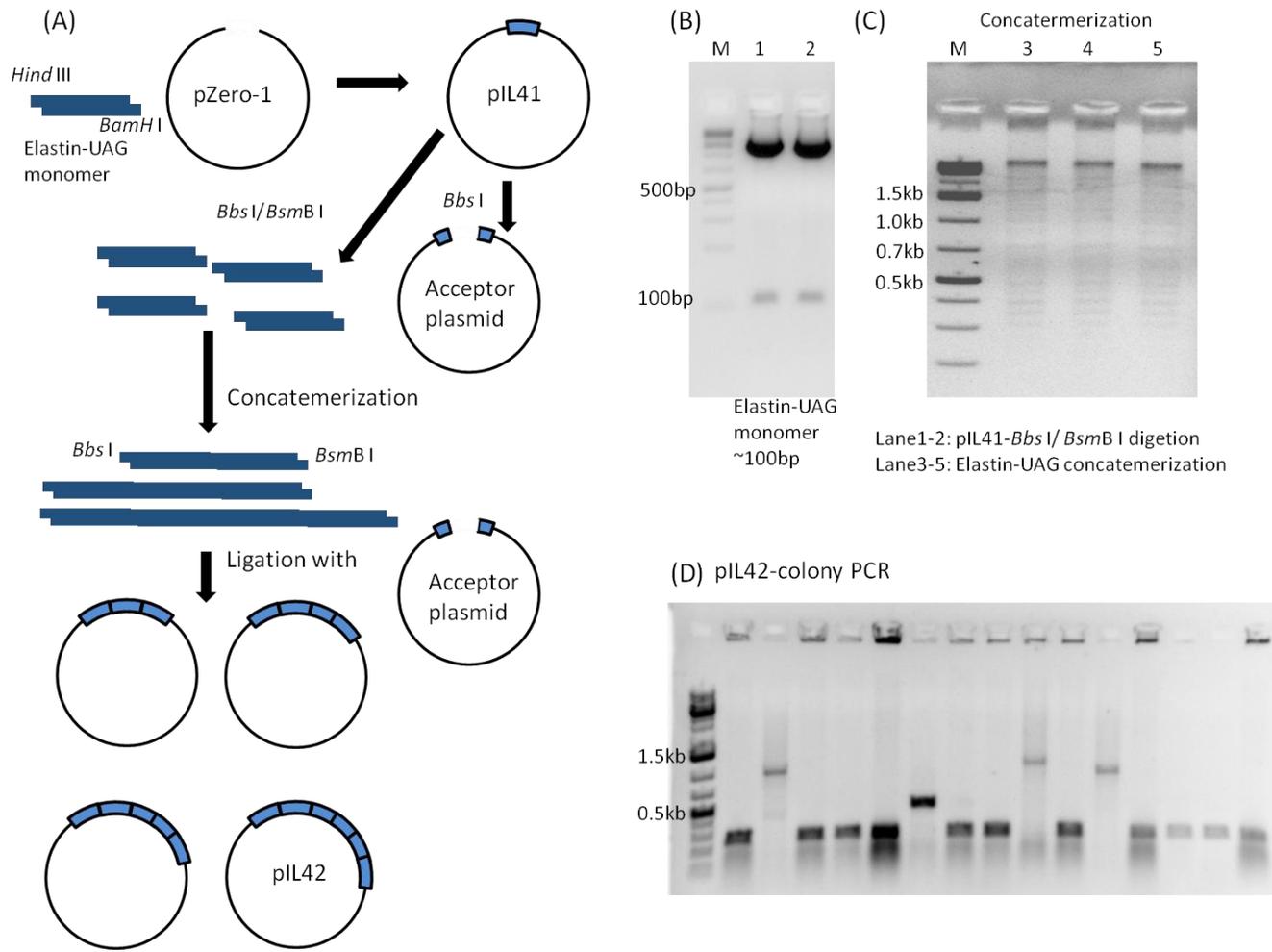


Figure 9. Construction of the Elastin-UAG. (A) Scheme of the synthesis of repetitive elastin-mimetic polypeptides. DNA gel electrophoresis of (B) digestion of pIL41 with *Bbs*I / *Bsm*B I, (C) DNA concatemerization, (D) colony PCR of pIL42 recombinants. plasmids.

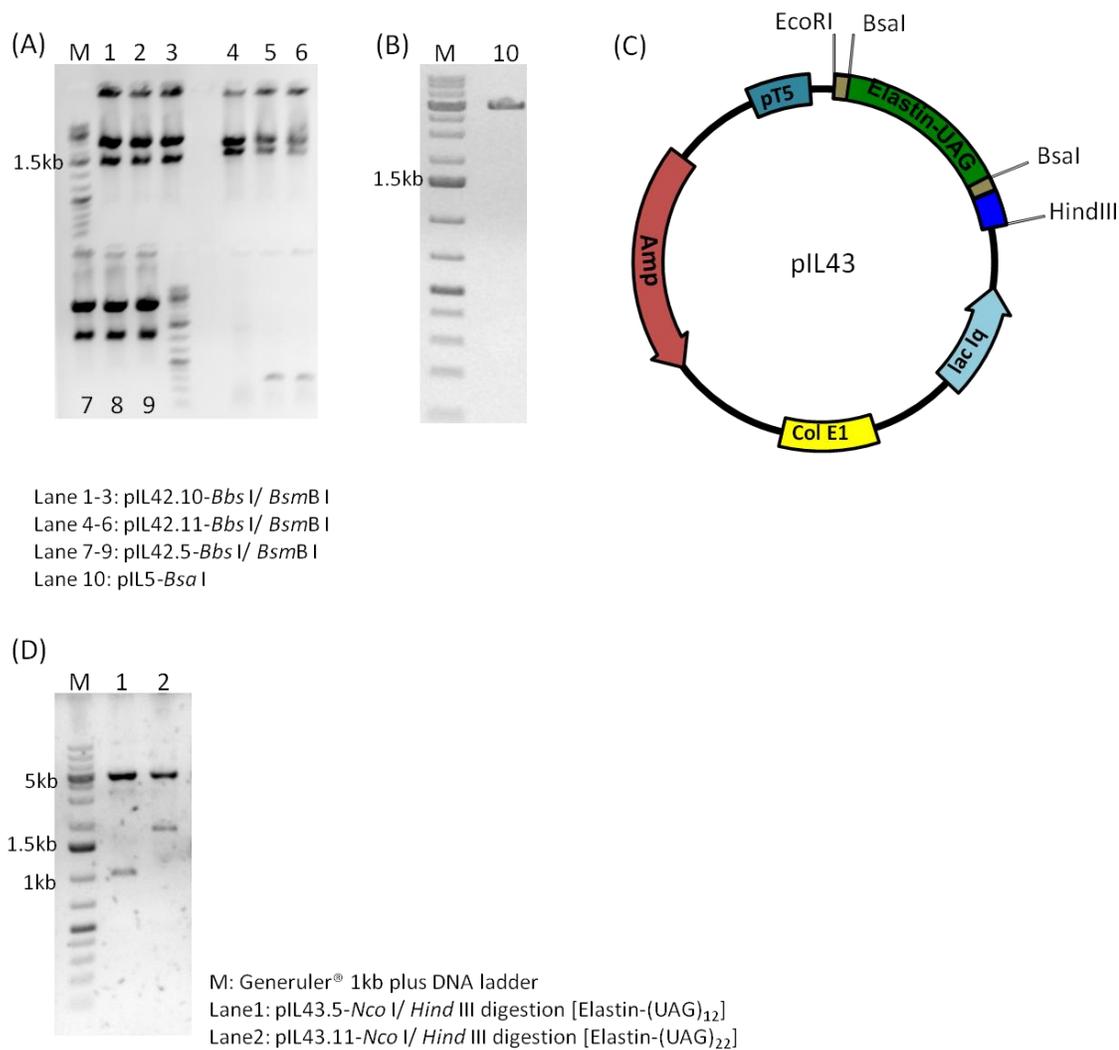


Figure 10. Construction of Elastin-UAG expression vectors. DNA gel electrophoresis of (A) digestion of pIL42-multimers with *Bbs I* / *BsmB I*, (B) digestion of pIL5 with *Bsa I*. (C) Plasmid **pIL43** encoding elastin-UAG DNA concatemers in pIL5. (D) DNA gel electrophoresis of the digestion of **pIL43.5** [Elastin-(UAG)₁₂] and **pIL43.11** [Elastin-(UAG)₂₂] with *Nco I* / *Hind III*.

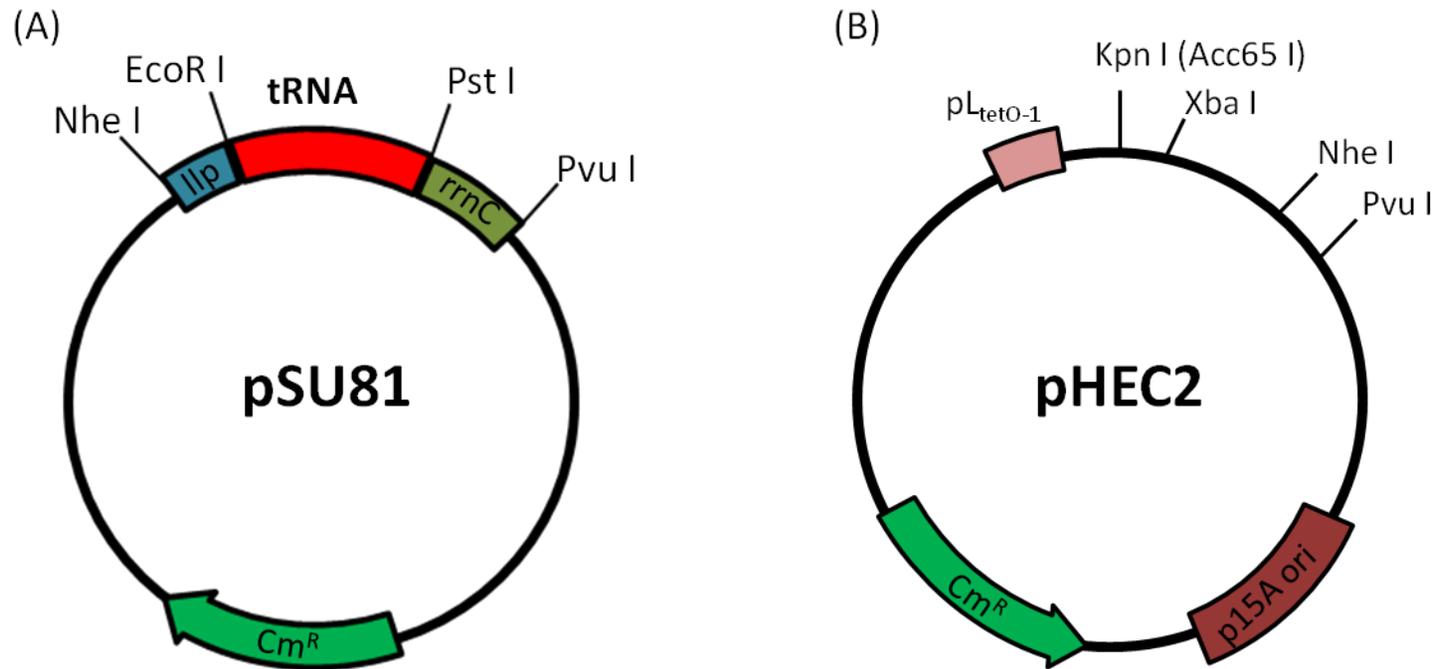


Figure 11. Plasmid map of pSU81 (A) and pHEC2 (B). Plasmid pSU81 is a useful plasmid for cloning orthogonal pairs as this plasmid enables cloning of tRNA genes as a cassette with flanking tRNA promoter and terminator sequences for optimal and constitutive expression of tRNA genes in an *E. coli* host³⁶. The pHEC2 plasmid is derived from pPROTet/Lar plasmid. The pHEC2 plasmid was generated by inverse PCR amplification with primers design to incorporate the *Nhe* I and *Pvu* I restriction endonuclease sites into the plasmid to enable cloning of tRNA cassettes in the plasmid³⁴. The multiple cloning site in the plasmid also contains the *Kpn* I and *Xba* I restriction endonuclease sites for cloning aaRS genes into the plasmid. The plasmid is a highly useful vector for cloning and expression of an orthogonal aaRS/tRNA pair in *E. coli*.

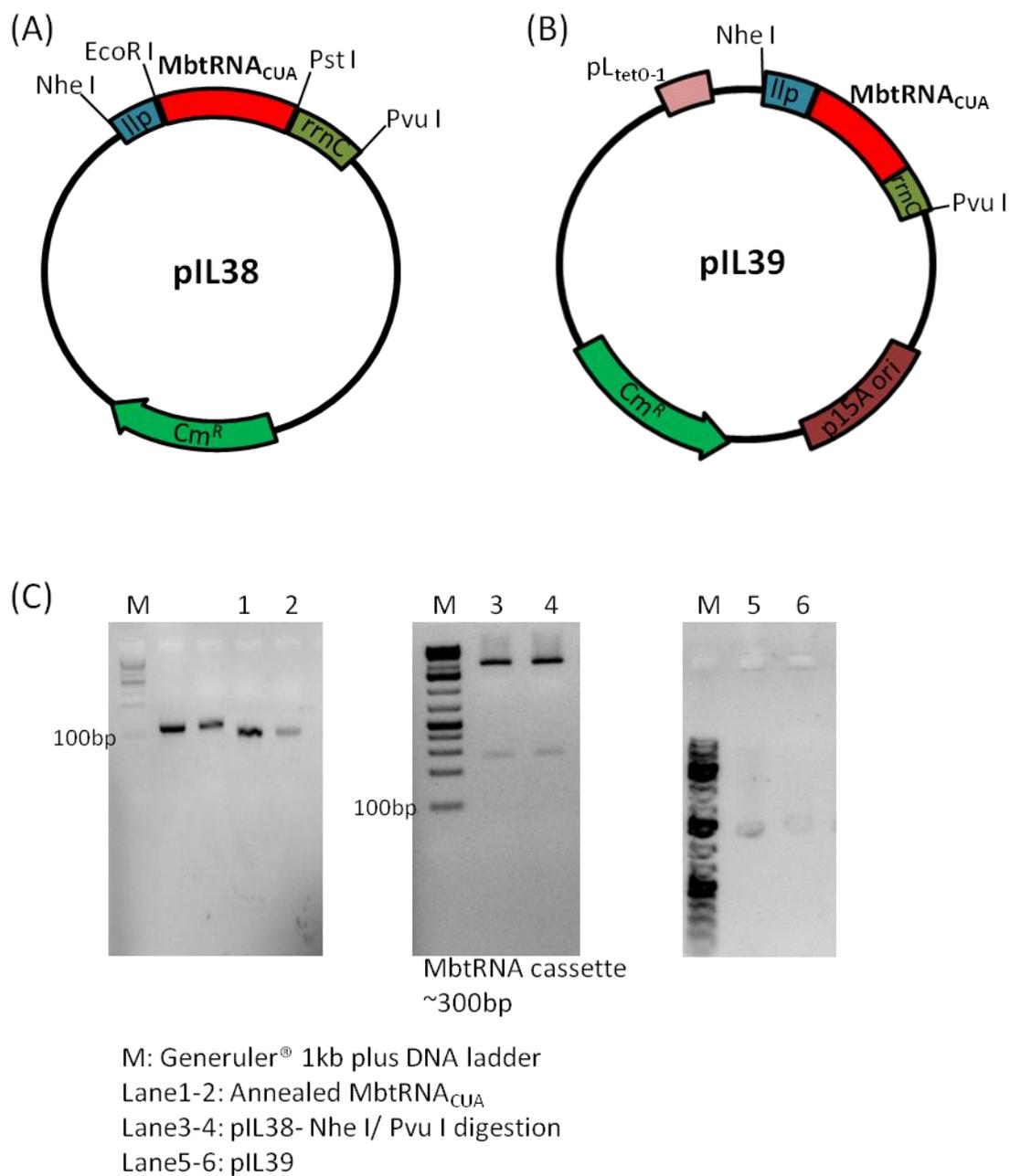


Figure 12. Plasmid pIL38 (A) and pIL39 (B) encoding MbtRNA_{CUA} in pSU81 and pHEC2 respectively. (C) DNA agarose gel electrophoresis of annealed MbtRNA_{CUA}, pIL38, and pIL39.

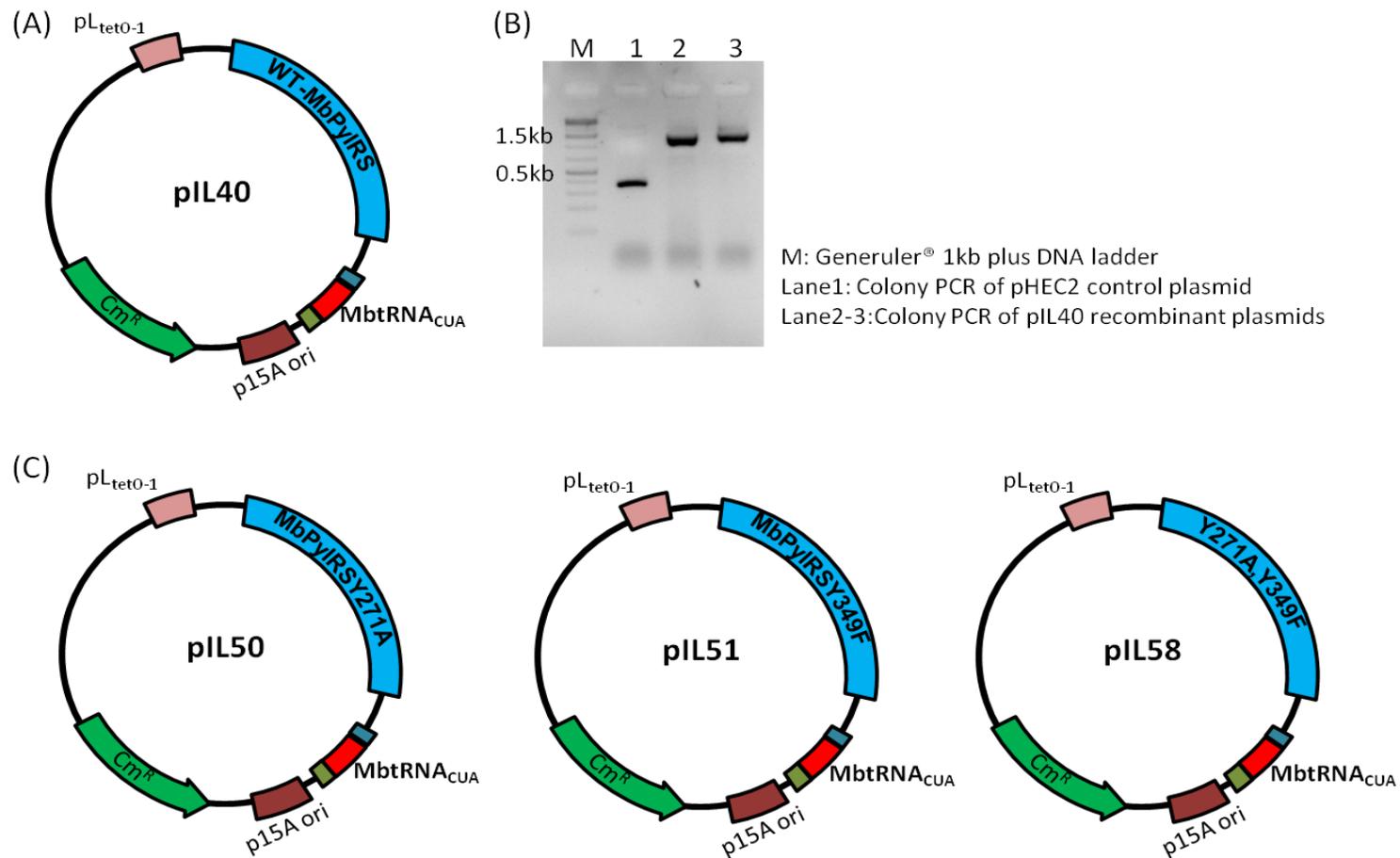


Figure 13. Plasmids encode MbPylRS pairs in pHEC2. (A) Plasmid pIL40 encodes wild-type MbPylRS. (B) PCR of *E.coli* colonies with desired recombinant plasmids. (C) Plasmid pIL50, pIL51, and pIL58 encode MbPylRS mutant pairs (Y271A, Y349F, Y271A/Y349F) respectively.

Table 1. Plasmids and strains utilized in Chapter 2.

Plasmids	Relevant characteristics	Reference
pQE80-L	<i>E. coli</i> cloning vector, Amp ^R	QIAGEN
pIL1	pQE80L with removal of internal <i>Nco</i> I, Amp ^R	This chapter
pIL2	pQE Adaptor in pIL1, Amp ^R	This chapter
pIL5	pIL2 with removal of internal <i>Bsa</i> I, Amp ^R	This chapter
pJ201-sfGFPCCC	sfGFP-CCC gene synthesized by DNA2.0, Kan ^R	This chapter
pJ201-sfGFPUAG	sfGFP-(UAG) ₄ gene synthesized by DNA2.0, Kan ^R	This chapter
pIL7	sfGFP-CCC in pIL2, Amp ^R	This chapter
pIL77	sfGFP-(UAG) ₄ in pIL2, Amp ^R	This chapter
pIL80	sfGFP-(UAG) ₃ in pIL2, Amp ^R	This chapter
pZero [®] -1	<i>E. coli</i> cloning vector, Zeocin ^R	Invitrogen
pIL41	Elastin-UAG monomer in pZero [®] -1, Zeo ^R	This chapter
pIL42.5	Elastin-UAG _{12mer} concatemer in pZero [®] -1, Zeo ^R	This chapter
pIL42.11	Elastin-UAG _{22mer} concatemer in pZero [®] -1, Zeo ^R	This chapter
pIL43.5	Elastin-UAG _{12mer} concatemer in pIL5, Amp ^R	This chapter
pIL43.11	Elastin-UAG _{22mer} concatemer in pIL5, Amp ^R	This chapter
pSU81	Synthetic plasmid containing a tRNA expression cassette, Cm ^R	Reference ³⁶
pHEC2	Modified pPROTet.E133/LarA.231, Cm ^R	Reference ³⁴
pIL38	MbtRNA ^{pyl} _{CUA} in pSU81, Cm ^R	This chapter
pIL39	MbtRNA ^{pyl} _{CUA} expression cassette in pHEC2, Cm ^R	This chapter
pJ201-MbPylRS	MbPylRS gene synthesized by DNA 2.0, Kan ^R	This chapter

pIL40	Wild-type MbPylRS pair in pHEC2, Cm ^R	This chapter
pIL50	MbPylRS (Y271A) in pHEC2, Cm ^R	This chapter
pIL51	MbPylRS (Y349F) in pHEC2, Cm ^R	This chapter
pIL58	MbPylRS (Y271A, Y349F) in pHEC2, Cm ^R	This chapter
Strain	Genotype	Reference
Top10F'	F [<i>lacIq</i> , Tn10(TetR)] <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) Φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139</i> Δ(<i>ara leu</i>) 7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i>	Invitrogen
DH10B	F' <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 endA1 araD139</i> Δ (<i>ara, leu</i>)7697 <i>galU galK λ- rpsL nupG</i> /pMON14272 / pMON7124	Invitrogen
MRA30	<i>prfA1 recA rph</i>	Reference ¹⁵

Table 2. Primers utilized in Chapter 2.

Name	Sequence (5'→3')
pQE-Adaptor-F	GCTTAGAATTCATTAAGAGGAGAAATTAACCATGGTTCCAAGAGACCGTCGACGGTCTCGTC CAGGTGTTGGCGGATCCGACGACG
pQE-Adaptor-R	GGACCAAGCTTATTAGTGATGATGATGATGATGATGATGATGATGGCCCTTGTCGTCGTCGTC GGATCCGCCAACACCTGGACGAGA
pQE80LbsaI-F	CAGTGCTGCAATGATACCGCGAGTGCCACGCTCACCGGCTCCAGATT
pQE80LbsaI-R	AATCTGGAGCCGGTGAGCGTGGCACTCGCGGTATCATTGCAGCACTG
sfGFP3TAG-F	GAGGAGAAATTAACCATGAGCAAGGGCGAAGAATTATTCACCGGCGTTCG
sfGFP3TAG-R	CGACGCCGGTGAATAATTCTTCGCCCTTGCTCATGGTTAATTTCTCCTC
Elastin-UAG-F	AGCTTGAAGACGTTCCAGGTGTTGGCGTACCGGGTGTAGGCGTTCGGGGTTAGGGTGTTCAG GCGTAGGTGTACCGGGTGTAGGTGTTCCAAGAGACGG
Elastin-UAG-R	GATCCCGTCTCTTGGAACACCTACACCCGGTACACCTACGCCTGGAACACCCTAACCCGGAAC GCCTACACCCGGTACGCCAACACCTGGAACGTCTTCA
pylTtRNA-F	AATTCGGAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTTCAGCCGGGTTAGATTCC CGGGGTTTCCGCCACTGCA
pylTtRNA-R	GTGGCGGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCATTCGATCTACATGA TCAGGTTTCCG

MbpylRS Y271A-F	GTCCGATGTTGGCACCGACCCTGGCTAACTACCTGCGCAAGCTGGACCG
MbpylRS Y271A-R	CGGTCCAGCTTGCGCAGGTAGTTAGCCAGGGTCGGTGCCAACATCGGAC
MbpylRS Y349F-F	CGTGGGCGATAGCTGCATGGTGTTCGGTGACACTCTGGACATCATGCAC
MbpylRS Y349F-R	GTGCATGATGTCCAGAGTGTCCACCGAACACCATGCAGCTATCGCCCACG

Results and Discussion

The most common method for site-specific incorporation of unnatural amino acids was based on the amber suppression strategy. While a number of biologically orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs have been developed to facilitate incorporation of single-site substitutions in response to amber codons in *E. coli*, we focused initially on the *M. barkeri fusaro* pyrrolysyl-tRNA synthetase (MbPylRS)/tRNA^{pyl}_{CUA}^{18,19} system for the incorporation of non-canonical amino acid substrates. A codon-optimized gene encoding MbPylRS was cloned into the vector under the transcription control of the *P_{Ltet}* promoter in which we had previously cloned the gene encoding tRNA^{pyl}_{CUA} under the control of the constitutive *lpp* promoter. Single mutant (Y349F), and double mutant (Y271A, Y349F) of the wild-type MbPylRS were characterized to facilitate the activation of *N*- ϵ -tBoc/Aloc-lysine, and *N*- ϵ -Cbz-lysine derivatives (Scheme 1), respectively, for co-translational insertion into polypeptide sequences in response to amber (UAG) codons with relatively high efficiency.

E. coli strain MRA30 (MG1655, *recA56 prfA1*) was employed as the bacterial host system for these experimental studies. This strain carries a conditionally lethal mutation in the *prfA* gene within its genome¹⁵, which encodes a structural variant of RF1 with attenuated translation termination activity. The phenotype caused by the mutation is temperature sensitivity (Ts) for growth at 42 °C and increased readthrough of the stopcodons UAA and UAG. Even within the permissive temperatures range (30-37 °C), the activity of the mutant RF1 is still weakened compared with the wild-type protein. For comparison, we employed *E. coli* DH10B, a common laboratory expression strain that

supports single-site insertion of non-canonical amino acids, as a control host in these studies as it presumably displays wild-type levels of RF1 activity.

We constructed two test substrates to assess the suppression efficiency of the RF1-attenuated *E. coli* host system and evaluate its suitability for the incorporation of non-canonical amino acids at multiple positions within a recombinant polypeptide sequence. The first genetic construct was derived from a variant of the green fluorescent protein superfolder (sfGFP)²⁹, in which three amber (UAG) codons were inserted at permissible loop positions within the coding sequence. This fluorescent reporter system permitted the use of flow cytometric analysis to monitor the incorporation of the non-canonical amino acids into the model protein. The second genetic construct was derived from an elastin-mimetic protein polymer^{38,39} based on a repeat sequence [Val-Pro-Gly-Xaa-Gly] in which the introduction of an amber codon was employed to specify the position and frequency of occurrence of a non-canonical amino acid (Xaa). This genetic template permitted evaluation of the potential of our approach for the preparation of sequence-repetitive polypeptides that had been substituted with multiple non-canonical amino acids.

Suppression Efficiency Assay

A codon-optimized gene encoding the superfolder variant of green fluorescent protein (sfGFP) was synthesized in which three amber (UAG) codons were inserted at permissible loop positions³⁰ (Gln157/Lys158, Glu172/Asp173, and Leu194/Leu195) within the coding sequence. The sfGFP-(UAG)₃ gene was cloned into a modified pQE-80L expression plasmid pIL2. The resulting plasmid pIL80 is under control of the phage T5lac promoter, which supports inducible expression in response to the presence of

inducer IPTG. In the presence of MbPylRS pairs, non-sense suppression with non-canonical amino acids at the three amber codon positions affords the expression of full-length sfGFP variant, which can be detected via its fluorescent emission. Fluorescence-activated cell sorting (FACS) analysis was employed to monitor incorporation of *N*- ϵ -alkoxycarbonyllysine derivatives into the sfGFP-(UAG)₃ reporter in *E. coli* strains DH10B and MRA30. The FACS data clearly indicated an increase in the fluorescent activity of the sfGFP-(UAG)₃ derivatives in *E. coli* strain MRA30 in the presence of the pyrrolysine analogues with respect to the negative control (absence of analogue) under conditions of the co-expression with the mutant MbPylRS pair (Fig. 14). In comparison, the fluorescent activity under the identical conditions for *E. coli* strain DH10B cannot be effectively distinguished from the negative control. The FACS data demonstrate that *E. coli* strain MRA30 supports a higher level of full-length sfGFP expression in the presence of the non-canonical amino acids, which suggests the suppressor tRNA competes more effectively for amino acid insertion due to the attenuated activity of the mutant RF1. The fluorescent signal for the sfGFP derivatives showed varying degrees for the three analogues, which agrees with the previous reports of their incorporation efficiency in single-site suppression studies.^{24,25} The observed differences may due to several causes, including the efficiency of substrate recognition and tRNA charging, the binding efficiency of the charged tRNA to EF-Tu and the ribosome, residual competition between the aminoacylated suppressor tRNA and the mutant release factor.

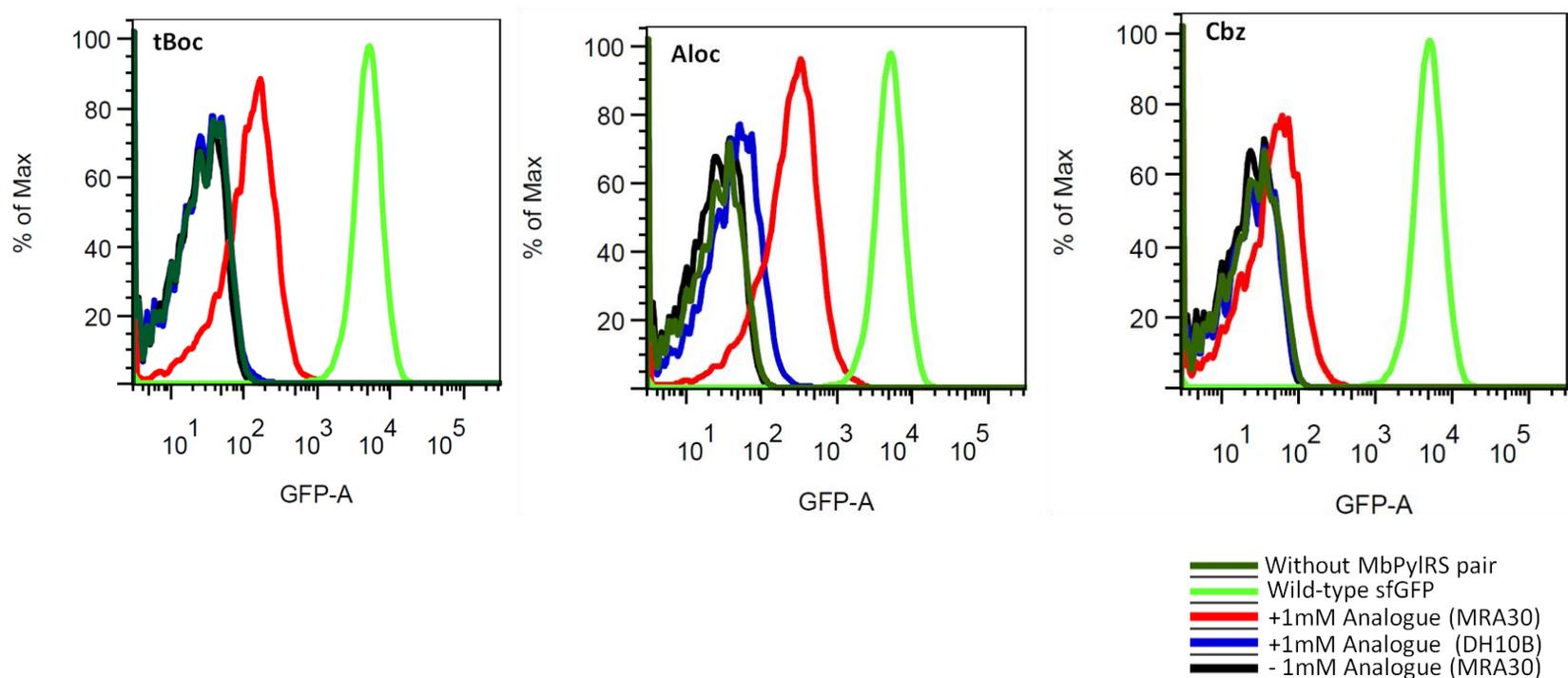


Figure 14. FACS analysis of sfGFP derivatives for the indicated non-canonical amino acid. (Legend: light green, *E. coli* MRA30 with wild-type sfGFP (positive control); red, *E. coli* MRA30 with the co-expression of MbPylRS pair in the presence of non-canonical amino acid; blue, *E. coli* DH10B with the co-expression of MbPylRS pair in the presence of non-canonical amino acid; dark green, *E. coli* MRA30 without the co-expression of the MbPylRS pair (negative control); and black, *E. coli* MRA30 with the co-expression of MbPylRS pair and without non-canonical amino acid (negative control)).

Characterization of sfGFP derivatives

In order to quantify the effectiveness of the suppression for the various analogues, sfGFP derivatives in the presence of the respective non-canonical amino acids were expressed and purified from *E. coli* strain MRA30. Western blot analyses of whole-cell lysates indicated a significantly higher level of expression of the target protein in *E. coli* strain MRA30 when compared to the expression under identical conditions in *E. coli* strain DH10B (Fig. 15). The sfGFP derivatives were isolated from LB expression cultures supplemented with the pyrrolysine analogues using immobilized metal affinity chromatography and the yield was determined using the Bradford assay (Fig. 15 and Table 3). The yields differ between the non-canonical amino acid derivatives in a manner that reflects the trend observed in the FACS analysis of the sfGFP derivatives and western blot analyses. The sfGFP expression studies demonstrate that multiple termination codons could be decoded with a structurally diverse variety of non-canonical amino acids to afford chemically modified proteins in a good yield. To further characterize the incorporation of these non-canonical amino acids, mass spectrometry analysis was performed on the purified sfGFP variants (Fig. 16). Deconvoluted ESI-mass spectra indicated that the main peak of the sfGFP-(Xaa)₃ was greater than that for the wild-type sfGFP. The mass increment in each case was consistent with the presence of three equivalents of the non-canonical amino acid that had been inserted into the coding sequence of sfGFP (Table 3).

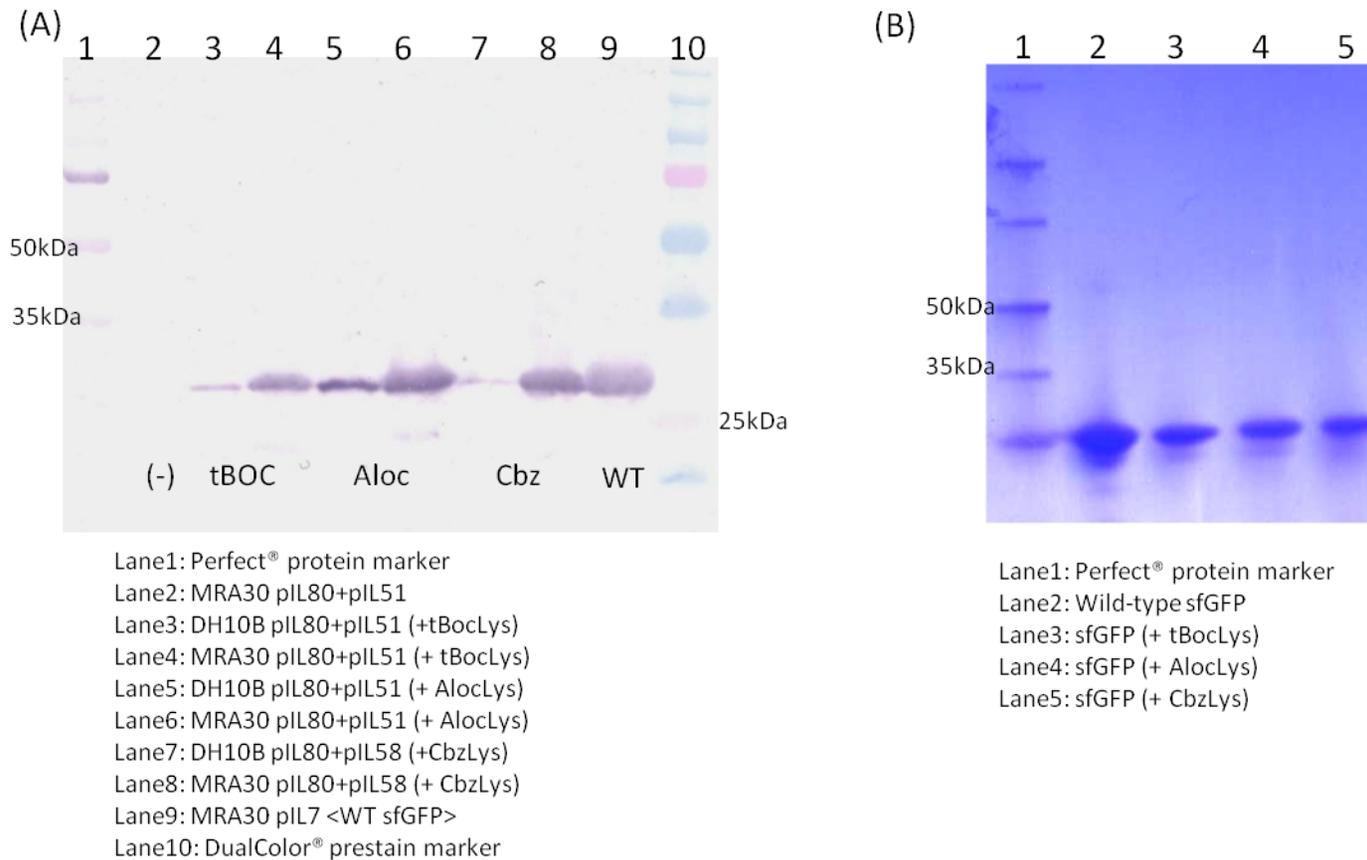


Figure 15. Expression studies of sfGFP derivatives modified with pyrrolysine analogues. (A) Western blot analysis of the expression of sfGFP derivatives with pyrrolysine analogues in MAR30 and DH10B strains. (B) SDS-PAGE analysis of the purified sfGFP variants.

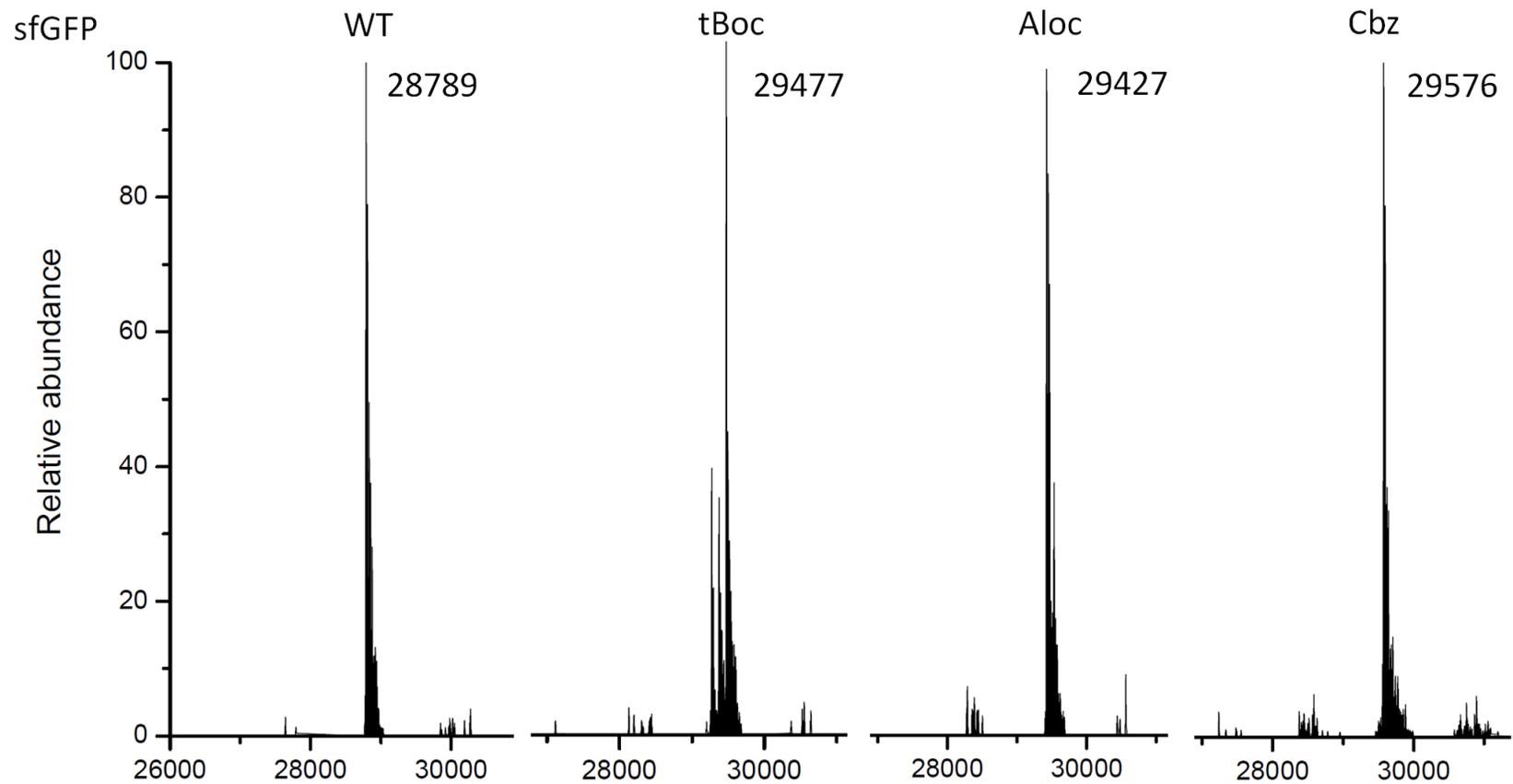


Figure 16. Deconvoluted ESI-mass spectra of the purified sfGFP derivatives from the expression cultures of *E. coli* strain MRA30 in the presence of the respective MbPylRS mutant pair and pyrrolysine analogues. The masses are consistent with the insertion of the indicated non-canonical amino acids at amber codon sites encoded within the sfGFP.

Table 3. Superfolder GFP expression and ESI-mass spectrometry data.

sfGFP	Yield (mg/L)^[a]	Theoretical m/z (u)^[b]	Experimental m/z (u)	(ΔWT)_{exp} (u)^[c]	(ΔWT)_{calc}(u)^[d]
WT	60.8	28791.3	28789	-	-
Aloc	20.5	29428.1	29427	638	637
tBoc	21.0	29476.2	29477	688	685
Cbz	5.5	29578.2	29576	787	787

[a] Yield was determined by the Bradford assay on purified protein solutions.

[b] Theoretical molar mass was calculated based on the assumption that *E. coli* methionyl-aminopeptidase results in proteolytic cleavage of the *N*-terminal methionine to afford the (sfGFP–Met) derivatives.

[c] The difference in experimental molar masses between the substituted sfGFP-(Xaa)₃ and wild-type sfGFP.

[d] Calculated molar mass for three equivalents of the indicated non-canonical amino acid as a peptidic unit (calcd molar mass: Aloc, C₁₀H₁₆N₂O₃, 212.25 u; tBoc, C₁₁H₂₀N₂O₃, 228.29 u; Cbz, C₁₄H₁₈N₂O₃, 262.31 u.)

Multi-site suppression of sequence-repetitive polypeptides

Synthesis of the Elastin-UAG genes

Multiple-site specific incorporation of unnatural amino acids is required for the synthesis of modified protein polymers which usually contain many suppression sites due to the repetitive sequence. We assessed the suppression efficiency of the RF1-attenuated *E. coli* system using a test substrate based on a sequence-repetitive polypeptide derived from an elastin-mimetic repeat sequence^{40,41}. The macromolecular properties of the elastomeric domains can be emulated by synthetic polypeptides that are composed of a concatenated sequence of native oligopeptide repeat motifs. In addition, elastin-mimetic polypeptides display a well-defined correlation between repeat sequence and macromolecular properties, which enables the creation of a wide variety of synthetic elastin analogues with tailorable physical properties³⁸. As such, elastin-mimetic polypeptides represent the best-characterized biomimetic protein-based materials that have been prepared and analyzed to date^{42,43}.

The elastin-mimetic repeat sequence, [(Val-Pro-Gly-Val-Gly)₂(Val-Pro-Gly-Xaa-Gly)(Val-Pro-Gly-Val-Gly)₂] **elastin-UAG**, was used as the monomer in which the introduction of an amber (UAG) codon was employed to specify the position of a non-canonical amino acid (Xaa). Previous studies have demonstrated that the elastin-mimetic repeat sequence is tolerant of non-conservative substitutions at the fourth position of the pentapeptide unit, which makes this position an ideal site for introduction of non-canonical amino acids³⁸. Annealing of the elastin-UAG synthetic oligonucleotides assembled the duplex DNA monomer. The presence of the *Hind* III and *Bam*H I restriction sites at the 5' and 3' overhang, facilitates the insertion of the elastin-UAG

monomer into the multiple cloning region of the plasmid pZErO[®]-1. A seamless cloning procedure was employed for head-to-tail self-ligation of the DNA cassette to afford a pool of DNA concatemers^{4,44}. The library of concatemers was fractionated via agarose gel electrophoresis to generate a ladder of bands where the size of individual concatemers differs by the unit of a single DNA monomer. Concatemers in the desired size range of 1000 to 2500 bp were purified and re-cloned into the original elastin-UAG monomer acceptor plasmid. The recombinant clones were screened for the concatemer inserts of desired size. The concatemers with twelve [elastin-(UAG)₁₂], and twenty-two [elastin-(UAG)₂₂] repeats of the DNA monomer were isolated and cloned in to the pQE-80L modified plasmid pIL5 for expression studies (Fig. 10).

Expression of Elastin-UAG derivatives

The elastin-mimetic polypeptides, elastin-UAGs, were used as artificial polymers for biosynthetic incorporation of unnatural amino acids adapting amber suppression to incorporate unnatural amino acids periodically along the polypeptide chain. The plasmids were transformed into *E. coli* strains DH10B and MRA30 in conjunction with the respective MbPylRS pairs, to investigate the influence of the relative level of RF1 activity on the yield of the full-length target polypeptide. SDS-PAGE analysis of the whole-cell lysates from the expression cultures was employed to evaluate the production of the two elastin-mimetic polypeptides under IPTG induction in the presence of the non-canonical amino acids. A significant level of protein accumulation was observed in the *E. coli* MRA30, and only in the presence of the respective MbPylRS pair that was specific for the corresponding non-canonical amino acid. In contrast, *E. coli* DH10B strain did not display significant levels of protein production under identical conditions, nor was

protein expression observed in the *E. coli* MRA30 strain upon induction in the absence of the non-canonical amino acid (Fig. 17). The substituted elastin-mimetic polypeptides could be purified from the cell lysate using inverse temperature transition cycling (4 °C/30 °C).^{45,46} SDS-PAGE analysis of the purified polypeptides indicated that the bands corresponding to the purified elastin derivatives migrated at the correct mass relative to the protein standards. The elastin-(CbzLys)₂₂ polypeptide could not be isolated using either temperature cycling or metal affinity purification protocol. This may be due to substantial degradation of the polypeptide occurred during the isolation. However, western blot analysis provided strong evidence for the initial production of the corresponding full-length polypeptide (Fig. 17). The isolated yields of the elastin derivatives reflected a similar trend observed in the sfGFP derivatives in that the *N*- ϵ -tBoc-lysine derivative afforded the highest yield, while the *N*- ϵ -Cbz-lysine derivative provided the lowest yield (Table 4). The discrepancies in protein yield observed between pyrrolysine analogues presumably reflect the considerations described above for the corresponding sfGFP derivatives. These considerations include differences in the enzymatic activity of the respective MbPylRS mutants for the non-canonical amino acids, and in suppression efficiency of the charged suppressor tRNAs during ribosomal chain elongation. However, the expression studies indicate that the system could allow substantial levels of the recombinant protein expression that encode amber codons up to twenty within the single polypeptide chain.

Temperature turbidity experiments were conducted on the elastin-(Xaa)₁₂ pyrrolysine derivatives and monitored by UV-visible spectrometry. The thermal transitions of the polymers were observed by an increase in the absorbance. Cooling and

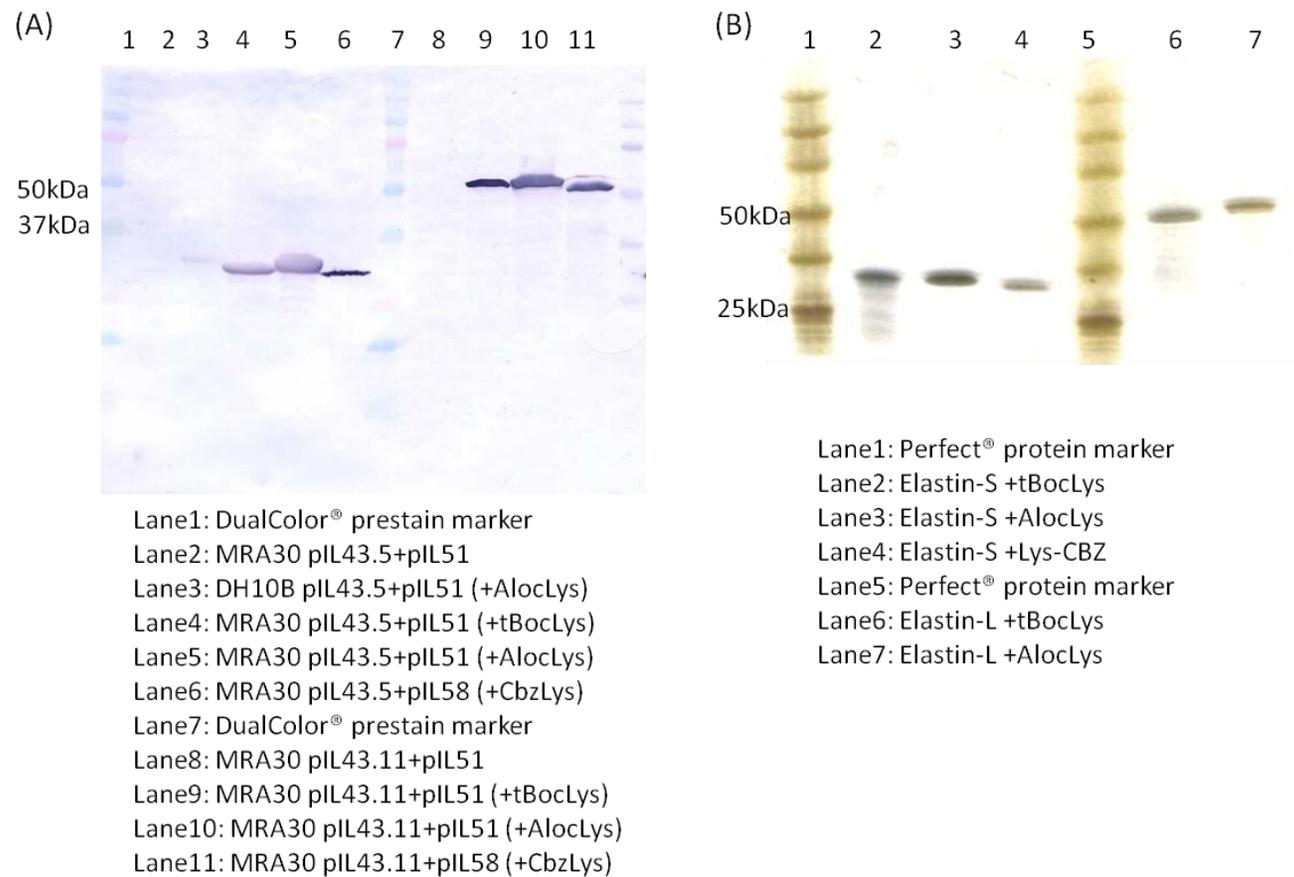


Figure 17. Expression of elastin-(UAG)_n derivatives with pyrrolysine analogues. (A) Western blot analysis of whole cell lysates of the elastin expression cultures with indicated analogues. Lane 2-6: elastin-(UAG)₁₂; Lane 8-11: elastin-(UAG)₂₂ (B) SDS-PAGE analysis of the purified elastin-mimetic polypeptides.

rescan of the samples *in situ* demonstrated the reversibility of the transitions (data not shown). These data indicated that temperature-driven transition of the elastin-(Xaa)₁₂ polymers occurred at a temperature approximately between 20 °C to 30 °C (Figure. 18). The temperature-dependent turbidimetry profile of these polymers suggests that the temperature-responsive aggregation occurred over a narrow temperature range. The elastin-(tBocLys)₁₂ polymers undergo a reversible coacervation characterized by a sharp transition at approximately 30 °C. The inverse temperature (T_t) of the elastin-(AlocLys)₁₂ polymers is calculated at approximately 29 °C, however, the aggregate formation might undergo over a broader temperature range compared to the tBoc derivatives. Furthermore, the transition of the elastin-(CbzLys)₁₂ polymer occurred at around 21 °C which is lower than tBoc- and Aloc-substitutions. The shift in the transition behavior of the different elastin-derivatives is most likely due to the hydrophobicity of the analogues in the fourth position of the pentapeptide⁴¹.

MALDI-TOF mass spectrometry

MALDI-TOF mass spectrometry was employed to characterize the full-length elastin-mimetic polypeptides that resulted from multi-site suppression at the encoded amber positions. While MALDI-mass spectra could be obtained for all of the elastin-(Xaa)₁₂ derivatives, only elastin-(AlocLys)₂₂ gave reliable data among the longer elastin derivatives. Presumably, this situation reflected the difficulties in the desorption and ionization of the relatively large, hydrophobic proteins or the concomitant gas-phase decomposition of the polypeptides in the mass spectrometer. The MALDI-TOF mass spectra in each case were consistent with a high level of substitution of the respective

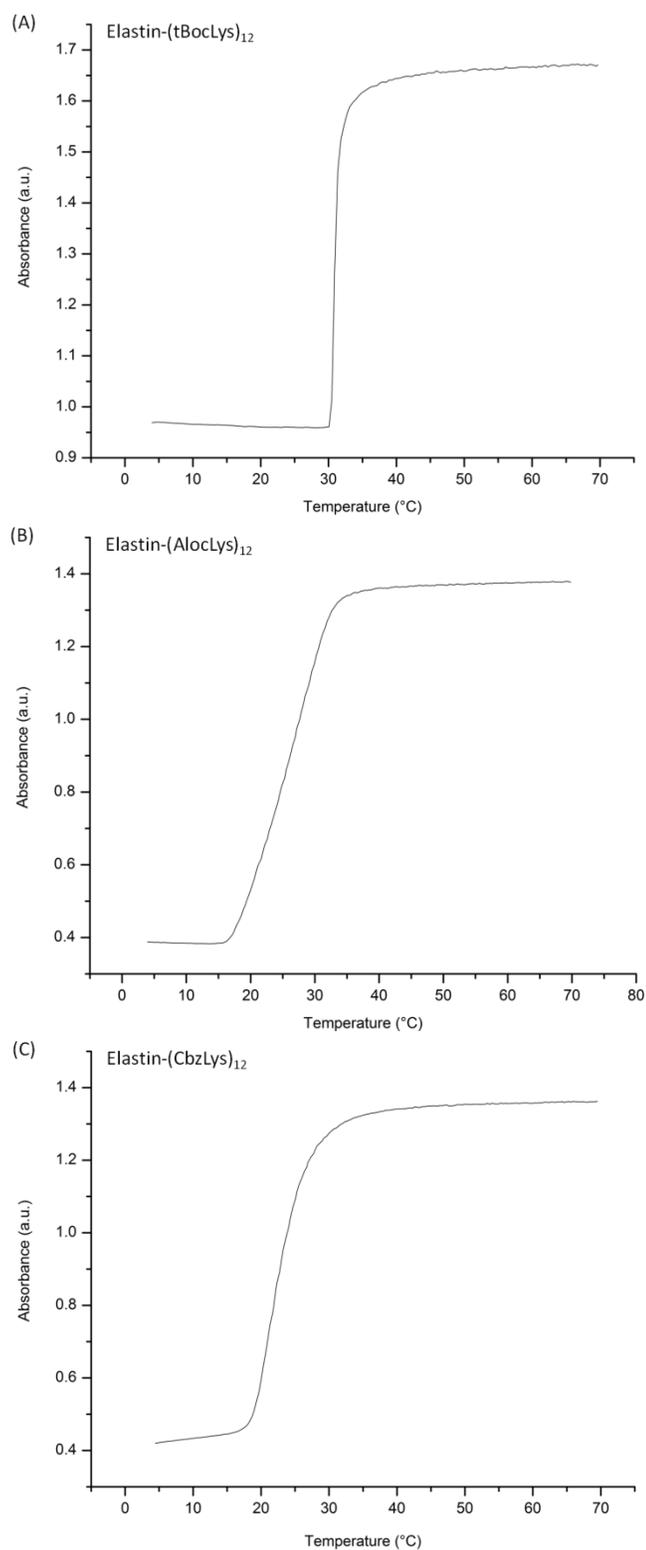


Figure 18. Temperature-dependent turbidimetry profiles of (A) the elastin-(tBocLys)₁₂, (B) elastin-(AlocLys)₁₂, and (C) elastin-(CbzLys)₁₂.

non-canonical amino acid into the corresponding elastin polypeptide sequence (Table 4). In addition to the major peak, the MALDI-TOF mass spectra of proteins isolated using inverse transition temperature cycling indicated the presence of a series of species at lower m/z values (Fig. 19), which might be due to the early termination at amber suppression sites or the extensive fragmentation during laser desorption. For the elastin-(AlocLys)_n derivatives ($n = 12, 22$), the difference between successive peaks corresponded to an average mass of 2143 units (u), which approximates the calculated mass of 2160 u for the [(Val-Pro-Gly-Val-Gly)₂(Val-Pro-Gly-AlocLys-Gly)(Val-Pro-Gly-Val-Gly)₂] repeat unit. The observed masses were consistent with truncation products that result from the termination at amber suppression sites upstream of the UAA codon that specifies the termination of the entire open reading frame of the elastin expression cassette. The truncation peaks are much more noticeable in the mass spectrum of elastin-(Aloc)₂₂, especially at lower m/z values. This situation may reflect the greater difficulty in achieving gas-phase ionization of the larger elastin derivative. These observations of the truncated products might suggest that the termination at amber sites can compete with the suppression. However, the SDS-PAGE analysis of the purified proteins indicates that even at the relatively high percentage of amber codons present within these elastin-mimetic polypeptides, full-length proteins are still obtained as the major products.

Tandem mass spectrometry

In order to ascertain the presence and the position of the non-canonical amino acid, tandem mass spectrometry was performed on cleavage fragments from the proteolytic digestion of the elastin-mimetic polypeptides. Due to the lack of convenient cleavage sites for the residue-specific endoproteases, such as trypsin, in the sequence-

repetitive protein polymer, metalloproteinase thermolysin was chosen for these analyses. Thermolysin catalyzes the selective hydrolysis of the peptide bonds at positions that are located directly upstream of the hydrophobic amino acid residues within the polypeptide sequence (Ile, Leu, Val, Ala, Met, Phe)⁴⁷ but the cleavage is impeded by the presence of a consecutive Pro residue. After thermolysin digestion, the high content of Val and Pro residues within the repeat units of the elastin-mimetic polypeptides produced several uniform cleavage fragments with detectable sizes for tandem mass spectrometry analysis. Two major fragments comprising [Val-Gly-Val-Pro-Gly-Xaa-Gly-Val-Pro-Gly] and [Val-Gly-Val-Pro-Gly-Xaa] in which the Xaa corresponds to the incorporated non-canonical amino acid are dominant after the protease digestion. Mass differences were observed between fragment ions that were attributed to the presence of *N*- ϵ -AlocLysine (212 u) (Fig. 20), *N*- ϵ -tBocLysine (228 u) (Fig. 21), and *N*- ϵ -CbzLysine (262 u) (Fig. 22). Mass spectrometric analyses of the thermolysin cleavage products derived from protein polymers corresponding to either 12 or 22 repeats of the elastin-mimetic sequences were similar in terms of the masses and the relative population of cleavage products. The fragments consistent with the incorporation of the respective non-canonical amino acids at the amber sites were selected for collision-induced dissociation to obtain the sequence information.

Tandem mass spectrometry was employed to analyze the fragments derived from the thermolysin cleavage of modified elastin polymers. The positive ESI-spectra of these digested peptides showed abundant $[M+H]^+$ and $[M+Na]^+$ ions. The protonated decapeptides, [VGVPGXGVPG], were selected for further fragmentation to obtain the sequence information. The ESI-MS/MS spectrum of $[VGVPG(AlocK)GVPG+H]^+$ at m/z

950 presented significant of b_8 and y_7 ions which demonstrated the easily cleavage at the N-terminus of proline residue that has been explained by the proline effect⁴⁸. The b_8 ion (m/z 778) showed the extensive neutral loss to form $b_8\text{-H}_2\text{O}$ (m/z 760), $b_8\text{-CO}$ (m/z 750), and $b_8\text{-C}_3\text{H}_5\text{OH}$ (m/z 720) which is the terminal modification on the ϵ -amino group of lysine (Fig. 20). In the case of the tBoc-lysine derivatives, the dominant peak was $[(\text{M}+\text{H})\text{-tBoc}+\text{H}]^+$ (m/z 866) where the tBoc group was lost in the gas phase and the observed mass difference (-100 u) corresponded to that of the deprotected lysine rather than the tBoc-lysine derivative (Fig. 21). The loss of the entire tBoc group from the lysine side chain could be explained by a McKafferty-type rearrangement⁴⁹ reported earlier which involves a γ -H migration from the *t*-butyl group to the carbonyl oxygen in the ϵ -amino group of lysine and followed by the loss of 2-methyl-prop-1-ene (C_4H_8 , 56Da) and the subsequent loss of CO_2 (44Da). The fragmentation of $[\text{VGVP}(\text{CbzK})\text{GVPG}+\text{H}]^+$ ions (m/z 1000) presented a similar pattern of the Aloc-peptides (Fig. 22). The b_8 and y_7 ions were the most abundant in the ESI-MS/MS spectra. The b_8 ion (m/z 828) also showed the extensive neutral loss to form $b_8\text{-H}_2\text{O}$ (m/z 810), $b_8\text{-CO}$ (m/z 800), and $b_8\text{-PhCH}_2\text{OH}$ (m/z 720). The fragmentation pattern derived from the collision-induced dissociation of selected parent ions provides the sequence information of corresponding peptides from the proteolytic cleavage products. The tandem mass data here provide strong evidence for the presence of the non-canonical amino acids in the appropriate position within the elastin-mimetic repeat unit.

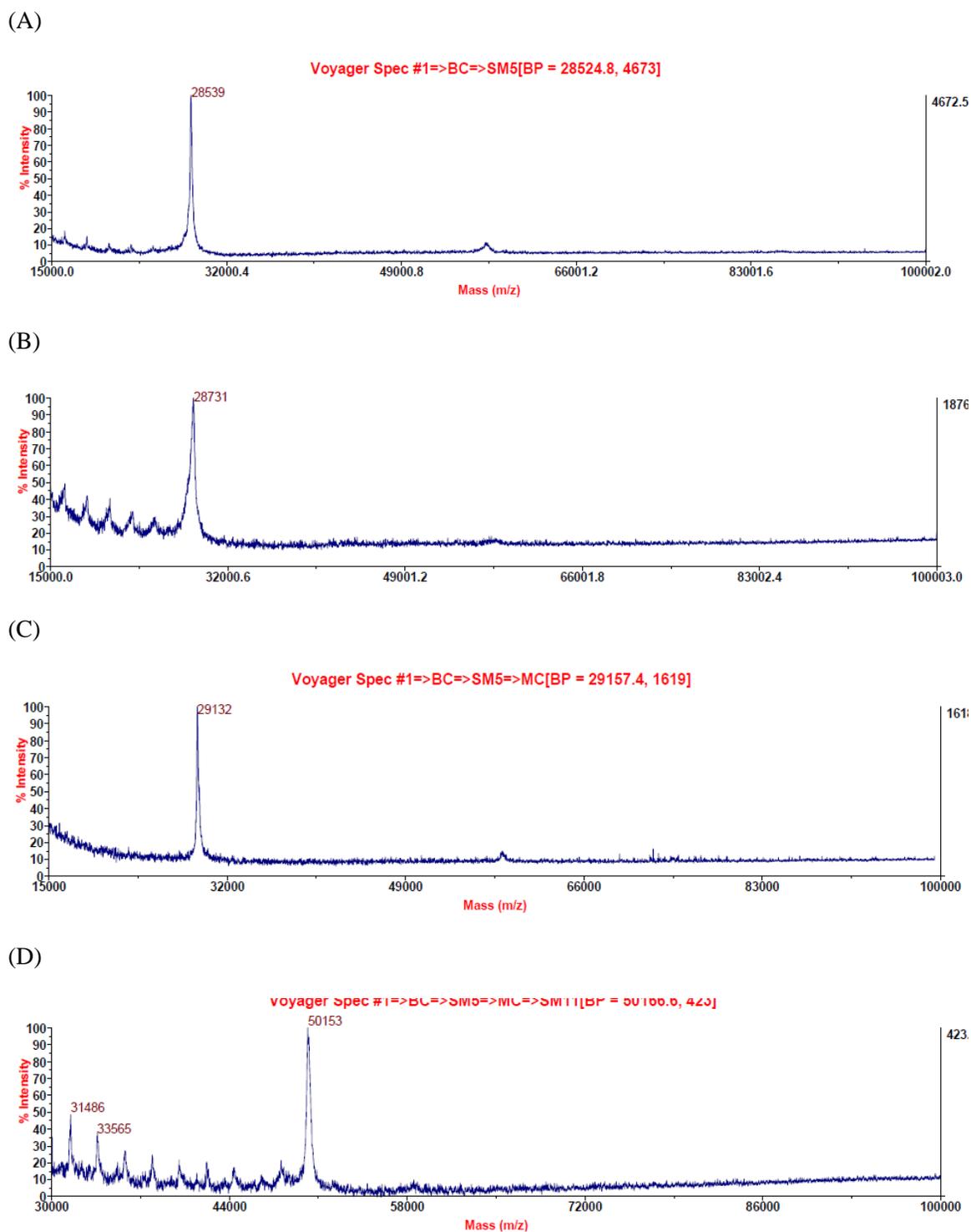


Figure 19. MALDI-TOF mass spectra of the full-length elastin derivatives: (A) elastin-(aloc)₁₂, (B) elastin-(tBoc)₁₂, (C) elastin-(CbzLys)₁₂, and (D) elastin-(alocLys)₂₂.

Table 4. Elastin protein expression and MALDI-TOF mass spectrometry data.

Elastin-(Xaa) _n	Yield (mg/L) ^[a]	Theoretical m/z ^{[b],[c]}	Experimental m/z
(Aloc) ₁₂	10	28540	28539
(Aloc) ₂₂	12	50147	50153
(tBoc) ₁₂	14	28733	28731
(tBoc) ₂₂	27	n.d. ^[d]	n.d. ^[d]
(Cbz) ₁₂	5.0	29141	29132

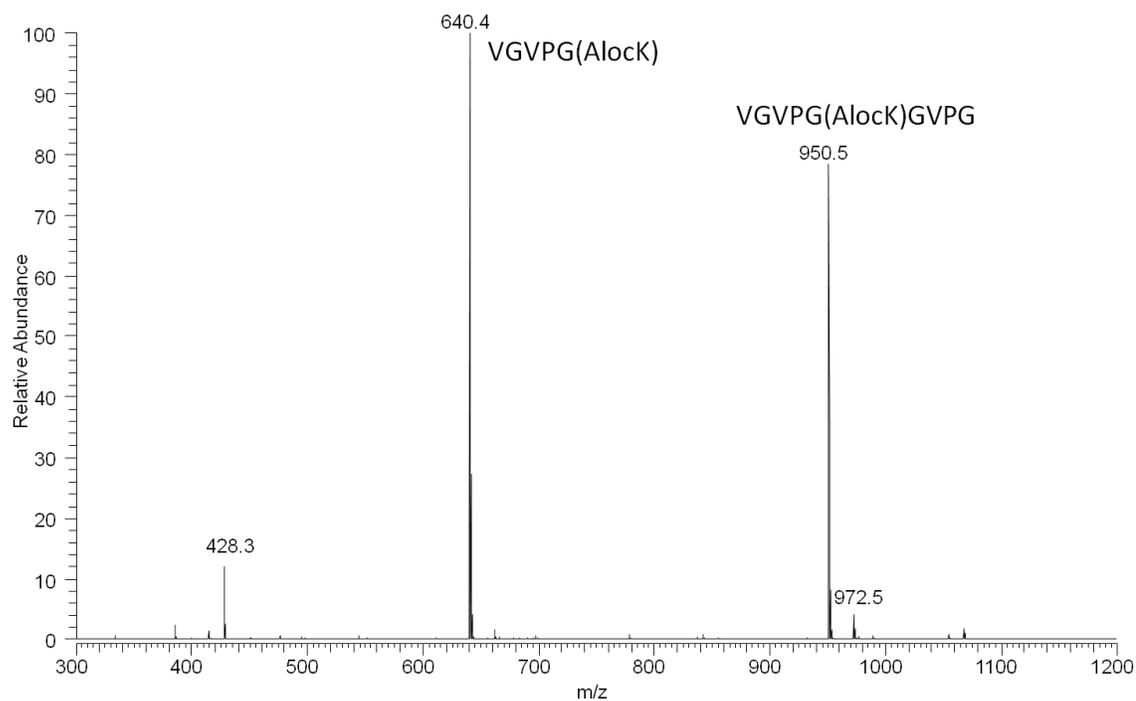
[a] Elastin proteins containing the *N*-ε-alkoxycarbonyl-lysine derivatives (Aloc, tBoc, and Cbz) were purified via inverse temperature cycling and the yield was measured by dry weight.

[b] Molar masses for the elastin derivatives were calculated based on the complete substitution with the non-canonical amino acid at the amber-encoded sites. The calculated molar mass assumes proteolytic cleavage of the *N*-terminal methionine residue as a consequence of the endogenous activity of the *E. coli* methionyl-aminopeptidase.

[c] Theoretical molar mass determinations were based on the amino acid sequence of the elastin polypeptides: [(VPGVG)₂VPGXG(VPGVG)₂]_nVPGVGGSSDDDDKGH₁₀ as [M+Na]⁺ species for the AlocLys, tBocLys, CbzLys derivatives.

[d] Not determined.

(A)



(B)

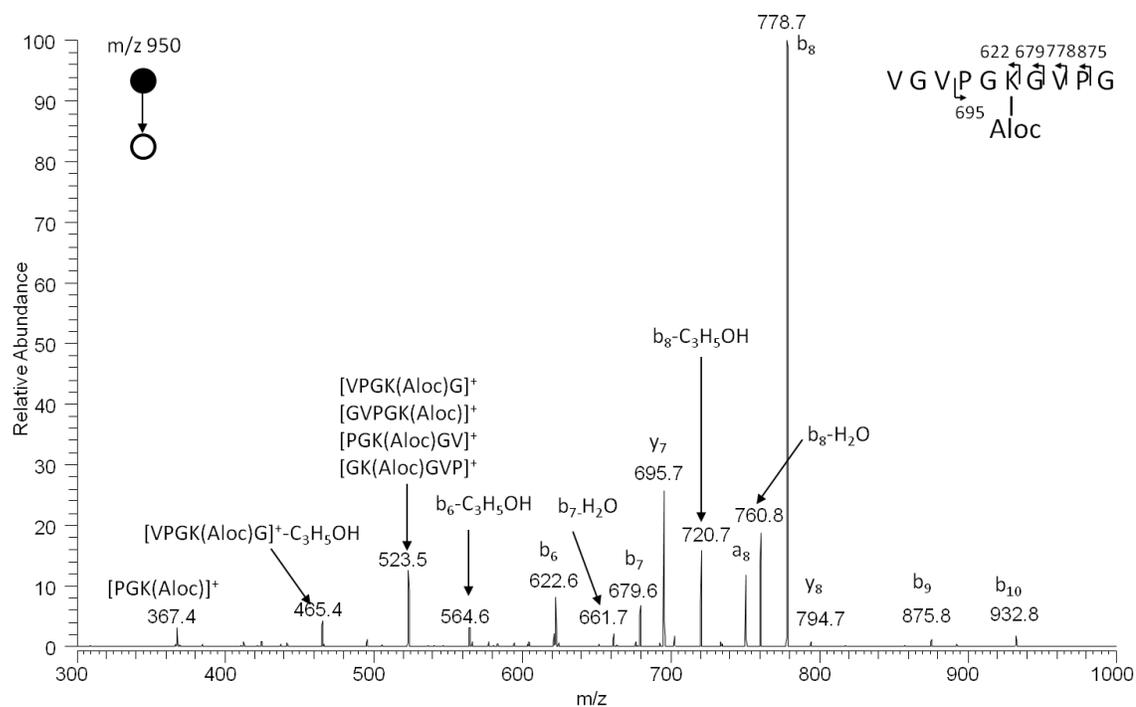
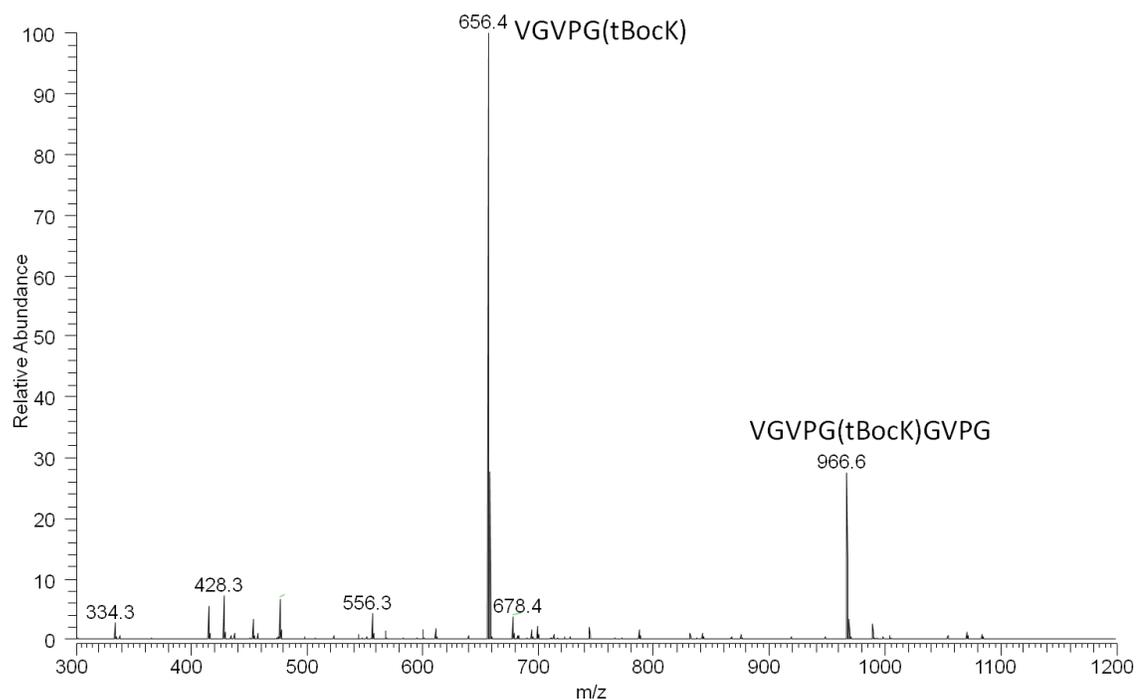


Figure 20. ESI-MS (A) and MS/MS (B) of elastin-(AlocLys)₁₂ derivatives digested with thermolysin.

(A)



(B)

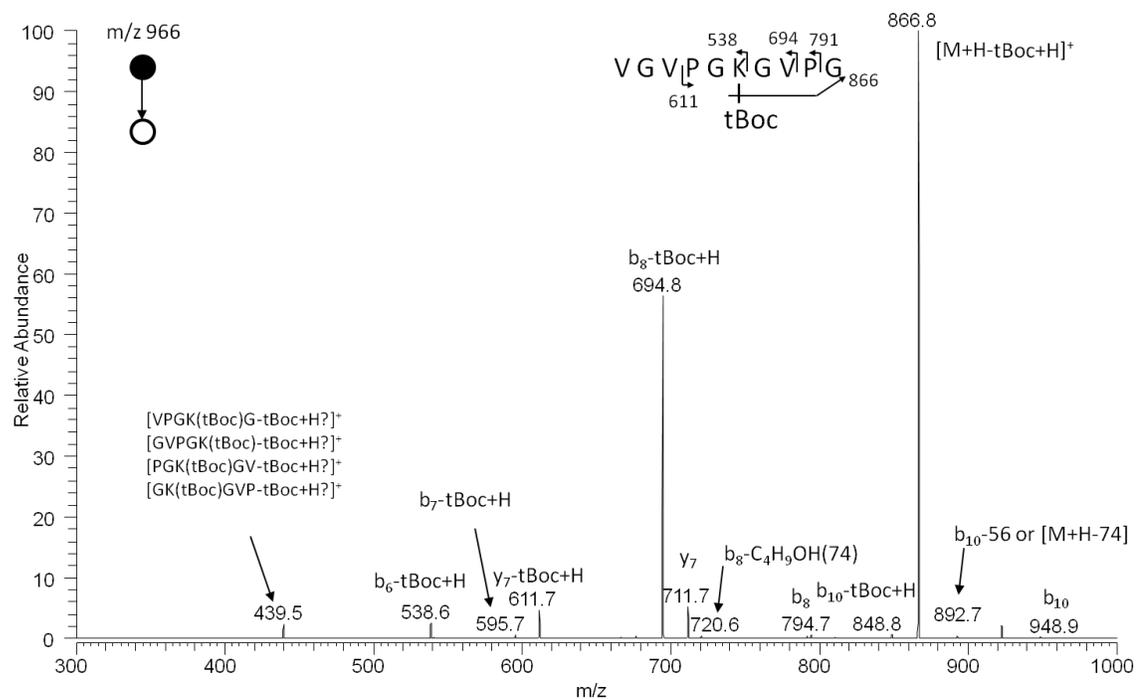
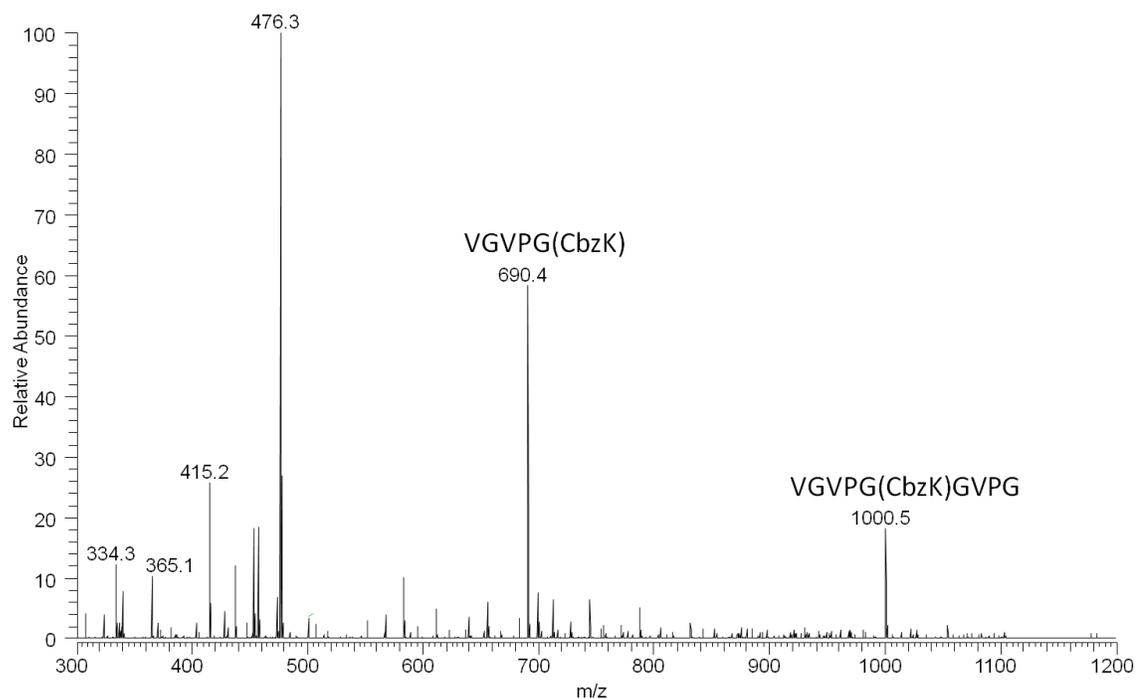


Figure 21. ESI-MS (A) and MS/MS (B) of elastin-(tBocLys)₁₂ derivatives digested with thermolysin.

(A)



(B)

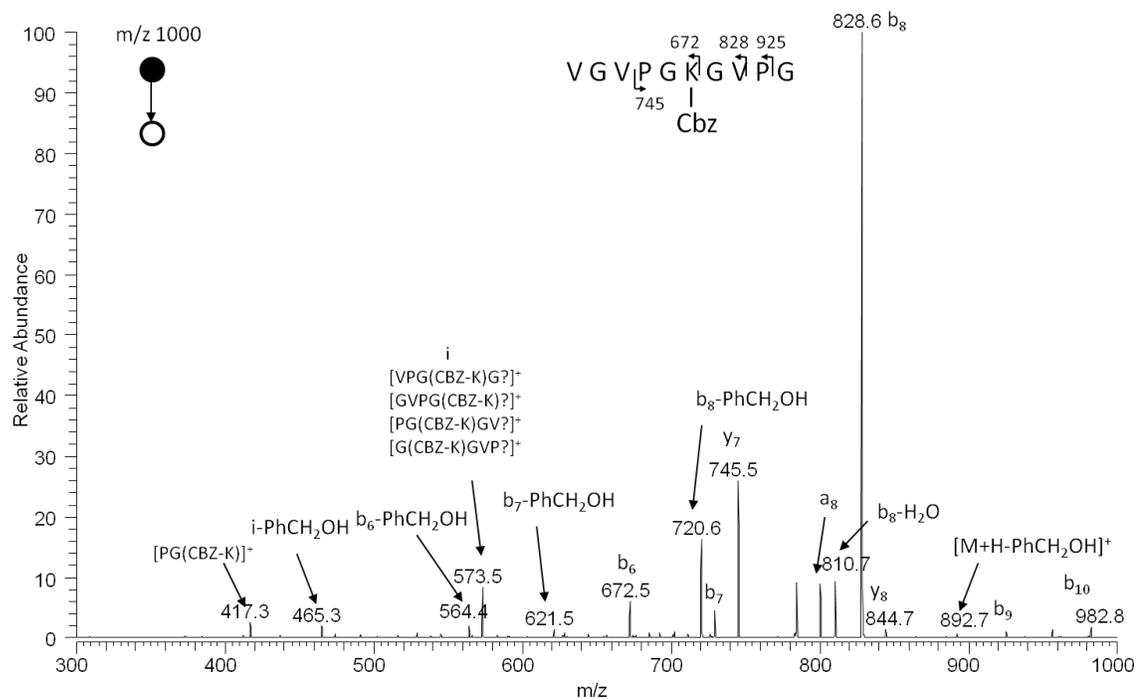


Figure 22. ESI-MS (A) and MS/MS (B) of elastin-(CbzLys)₁₂ derivatives digested with thermolysin.

Conclusion

In this chapter, we demonstrate a system for the multiple site-selective incorporation of non-canonical amino acids into the polypeptides based on the amber suppression strategy. A genetically engineered *E. coli* strain, MRA30, with an attenuated RF1 activity is selected as a host system with potentially enhanced suppression efficiency. The orthogonal aaRS/tRNA_{CUA} pair selected for the demonstration is derived from the *Methanosarcina barkeri* PylRS pair, which is naturally orthogonal in *E. coli* and comprises a native suppressor that decodes the amber codon. We constructed two model proteins to evaluate the suitability of our system for the insertion of unnatural amino acids at the multiple positions within a recombinant polypeptide sequence. Initially, a green fluorescent protein with triple non-sense insertions (sfGFP-(UAG)₃) is employed as a reporter. Successful suppression with the of the pyrrolysine analogues incorporated into sfGFP pairs retains its fluorescence, which can be monitored by FACS screen. In addition, mass spectrometry data of the purified sfGFP derivatives provide evidence that the *E. coli* strain employed in this study that characterized with an attenuated RF1 activity is competent for the multi-site insertion of amino acid analogues at the amber codons encoded within the corresponding genetic template.

Furthermore, this biosynthetic approach enables the preparation of novel protein polymers derived from the sequence-repetitive polypeptides where multiple non-canonical amino acids have installed at defined positions. We have demonstrated the expression of the elastin derivatives that have extensively modified with the pyrrolysine analogues in a significant yield, which implicates the application of this method for the design of functionalized protein polymers with selected amino acid analogues. In addition,

tandem mass spectrometry was employed to ensure the site-selective and the homogenous incorporation of unnatural amino acids by this approach. Our proposed system described herein facilitate the design of protein-based materials with expanded diversity of novel functionalities. Moreover, this work envisions a new scope in synthetic biology for investigating, exploring, and engineering proteins by incorporating selective amino acid analogues at multiple desired positions, for example, to mimic the protein post-translation modifications that are critical for the biological function. These advance provide a powerful set of new methodologies to generate proteins with enhanced or novel properties as well as new probes for the protein structural and functional studies.

Reference

- [1] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, 79, 413-444.
- [2] J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, D. A. Tirrell, *Curr. Opin. Chem. Biol.* **2010**, 14, 774-780.
- [3] T. S. Young, P. G. Schultz, *J. Biol. Chem.* **2010**, 285, 11039-11044.
- [4] R. A. McMillan, T. A. T. Lee, V. P. Conticello, *Macromolecules* **1999**, 32, 3643-3648.
- [5] Y. Qu, S. C. Payne, R. P. Apkarian, V. P. Conticello, *J. Am. Chem. Soc.* **2000**, 122, 5014-5015.
- [6] D. E. Meyer, A. Chilkoti, *Biomacromolecules* **2002**, 3, 357-367.
- [7] S. C. Payne, M. Patterson, V. P. Conticello, in *Protein Engineer Handbook, Vol. 2* (Eds.: S. Lutz, U. T. Bornscheuer), Wiley-VCH, **2009**, pp. 915-938.
- [8] L. Mi, *Biomacromolecules* **2006**, 7, 2099-2107.
- [9] A. Takasu, S. Kondo, A. Ito, Y. Furukawa, M. Higuchi, T. Kinoshita, I. Kwon, *Biomacromolecules* **2011**, 12, 3444-3452.
- [10] T. Mukai, A. Hayashi, F. Iraha, A. Sato, K. Ohtake, S. Yokoyama, K. Sakamoto, *Nucleic Acids Res.* **2010**, 38, 8188-8195.
- [11] T. Mukai, T. Yanagisawa, K. Ohtake, M. Wakamori, J. Adachi, N. Hino, A. Sato, T. Kobayashi, A. Hayashi, M. Shirouzu, T. Umehara, S. Yokoyama, K. Sakamoto, *Biochem. Biophys. Res. Commun.* **2011**, 411, 757-761.
- [12] D. B. Johnson, C. Wang, J. Xu, M. D. Schultz, R. J. Schmitz, J. R. Ecker, L. Wang, *ACS Chem. Biol.* **2012**, 7, 1337-1344.

- [13] D. B. Johnson, J. Xu, Z. Shen, J. K. Takimoto, M. D. Schultz, R. J. Schmitz, Z. Xiang, J. R. Ecker, S. P. Briggs, L. Wang, *Nat. Chem. Biol.* **2011**, *7*, 779-786.
- [14] S. M. Ryden, L. A. Isaksson, *Mol. Gen. Genet.* **1984**, *193*, 38-45.
- [15] S. Zhang, M. Ryden-Aulin, L. A. Kirsebom, L. A. Isaksson, *J. Mol. Biol.* **1994**, *242*, 614-618.
- [16] M. Kaczanowska, M. Ryden-Aulin, *J. Bacteriol.* **2004**, *186*, 3046-3055.
- [17] G. Srinivasan, *Science* **2002**, *296*, 1459-1462.
- [18] O. Namy, Y. Zhou, S. Gundllapalli, C. R. Polycarpo, A. Denise, J. P. Rousset, D. Soll, A. Ambrogelly, *FEBS Lett.* **2007**, *581*, 5282-5288.
- [19] D. G. Longstaff, R. C. Larue, J. E. Faust, A. Mahapatra, L. Zhang, K. B. Green-Church, J. A. Krzycki, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1021-1026.
- [20] T. Fekner, M. K. Chan, *Curr. Opin. Chem. Biol.* **2011**, *15*, 387-391.
- [21] S. K. Blight, R. C. Larue, A. Mahapatra, D. G. Longstaff, E. Chang, G. Zhao, P. T. Kang, K. B. Green-Church, M. K. Chan, J. A. Krzycki, *Nature* **2004**, *431*, 333-335.
- [22] C. Polycarpo, A. Ambrogelly, A. Berube, S. M. Winbush, J. A. McCloskey, P. F. Crain, J. L. Wood, D. Soll, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12450-12454.
- [23] A. Ambrogelly, S. Palioura, D. Soll, *Nat. Chem. Biol.* **2007**, *3*, 29-35.
- [24] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187-1197.
- [25] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2008**, *378*, 634-652.

- [26] J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Soll, T. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11268-11273.
- [27] S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044-1048.
- [28] T. H. Yoo, D. A. Tirrell, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 5340-5343.
- [29] J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **2006**, *24*, 79-88.
- [30] M. R. Abedi, G. Caponigro, A. Kamb, *Nucleic Acids Res.* **1998**, *26*, 623-630.
- [31] W. Kim, A. George, M. Evans, V. P. Conticello, *ChemBioChem* **2004**, *5*, 928-936.
- [32] W. Kim, R. A. McMillan, J. P. Snyder, V. P. Conticello, *J. Am. Chem. Soc.* **2005**, *127*, 18121-18132.
- [33] W. Kim, K. I. Hardcastle, V. P. Conticello, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 8141-8145.
- [34] H. E. Carpenter-Desai, Emory University **2008**.
- [35] M. A. Patterson, Emory University **2011**.
- [36] K. Gabriel, W. H. McClain, *J. Mol. Biol.* **2001**, *310*, 543-548.
- [37] R. Lutz, H. Bujard, *Nucleic Acids Res.* **1997**, *25*, 1203-1210.
- [38] D. W. Urry, C.-H. Luan, C. M. Harris, T. M. Parker, in *Protein-Based Materials* (Eds.: D. Kaplan, K. McGrath), Birkhauser, Boston, **1997**, pp. 133-177.
- [39] D. E. Meyer, A. Chilkoti, *Biomacromolecules* **2004**, *5*, 846-851.
- [40] D. W. Urry, C. H. Luan, T. M. Parker, D. C. Gowda, K. U. Prasad, M. C. Reid, A. Safavy, *J. Am. Chem. Soc.* **1991**, *113*, 4346-4348.

- [41] D. W. Urry, D. C. Gowda, T. M. Parker, C. H. Luan, M. C. Reid, C. M. Harris, A. Pattanaik, R. D. Harris, *Biopolymers* **1992**, *32*, 1243-1250.
- [42] V. P. Conticello, H. E. Carpenter Desai, in *Polymer Science: A Comprehensive Reference Vol. 9* (Eds.: K. Matyjaszewski , M. Möller), Elsevier, **2012**, pp. 71-103.
- [43] S. R. MacEwan, A. Chilkoti, *Biopolymers* **2010**, *94*, 60-77.
- [44] N. L. Goeden-Wood, V. P. Conticello, S. J. Muller, J. D. Keasling, *Biomacromolecules* **2002**, *3*, 874-879.
- [45] D. Y. Furgeson, M. R. Dreher, A. Chilkoti, *J. Control. Release* **2006**, *110*, 362-369.
- [46] D. T. McPherson, J. Xu, D. W. Urry, *Protein Expr. Purif.* **1996**, *7*, 51-57.
- [47] S. J. Bark, N. Muster, J. R. Yates, G. Siuzdak, *J. Am. Chem. Soc.* **2001**, *123*, 1774-1775.
- [48] C. Bleiholder, S. Suhai, A. G. Harrison, B. Paizs, *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1032-1039.
- [49] G. Raju, V. Ramesh, R. Srinivas, G. V. M. Sharma, B. S. Babu, *J. Mass Spectrom.* **2010**, *45*, 651-663.

Chapter 3

System optimization for the multi-site
specific incorporation of photo-crosslinkable
amino acid analogues

Introduction

Biological and mechanical responses of elastin-based biomaterials provide significant opportunities to modulate material microstructures that are tunable through the precise control of the primary amino acid sequence.¹⁻³ The native elastin network is generated through the post-translational modifications such as oxidative deamination of lysyl residues in the soluble precursor tropoelastin that results in the covalent crosslinking of polymer chains.⁴⁻⁶ In our lab, we have worked extensively to design and synthesize a new class of protein polymers that can form physically or chemically crosslinked elastin networks based on the consensus pentapeptide repeat motif, [Val-Pro-Gly-Xaa-Gly], derived from the native elastin.⁷⁻⁹ Generally, chemical or enzymatic crosslinking¹⁰ strategies that take advantage of the genetically encoded lysine residues at the Xaa positions are utilized to introduce the cross-links into the elastin network. Commonly used primary amine-reactive crosslinkers include isocyanates, phosphines, aldehydes, genipin, and N-hydroxysuccinimide (NHS) esters.¹¹ Although these approaches provide a measure of control over the degree of crosslinking, they might be slow, inefficient, and limited by the cytotoxicity of the chemical reagents. Additionally, the formation of the cross-links could result in unintended side reactions, byproducts, or residual amino groups, which may alter the responsive properties of the protein materials or create the non-crosslinked defect sites within the polymer network. In this work, a new class of elastin-mimetic polypeptides is designed and synthesized, in which the fourth position (Xaa) of the consensus pentapeptide repeat unit is selectively substituted with unnatural amino acid analogues. The non-canonical amino acids examined in this study possess the

ability to form covalent cross-links via photo-crosslinking could be employed as an alternative method to generate the elastin networks.

Photo-crosslinking is a method that has been successfully exploited for the identification and localization of macromolecular receptors within biological systems.¹²⁻¹⁵ Encoding the photo-crosslinker directly into the polypeptide chain allows the determination of protein interactions that are weak, transient, or pH-sensitive at distinct functional states. In addition, the activation of the photo-inducible probes can be exquisitely controlled. After the UV irradiation with the presence of the potential binding partners, those residues in proximity to the photo-crosslinker become covalently trapped, allowing the subsequent identification of factors that interact with the target protein. Benzophenone and phenylazido moieties represent a class of photo-activated probes, which have been utilized extensively for the study of protein-protein interactions and signaling protein complexes.¹⁶ The benzophenone class of photo-probes reacts preferentially with the unreactive C-H bonds of protein chains, even in the presence of solvent water and bulk nucleophiles.¹⁷ The excitation wavelength is in the near-UV at 350-360 nm, which avoids the damage of proteins and nucleic acids. Moreover, benzophenones do not photodissociate and their photo-activated triplet state readily relaxes in the absence of an unactivated C-H bond with which to react. This reversible excitation allows the repeated excitation by UV light without damage to the protein and thus results in a higher cross-linking yield.

Aryl-azides have also been used to photo-chemically label antibodies, selectively tag protein with probes, and map protein-protein interactions.^{18,19} In addition, the azido group can serve as IR-active probes²⁰ or possess unique chemical reactivity for further

modification. The advantages of azido chemistry are that the reaction is fast, viable at physiological conditions, and independent of pH over a wide range. The conjugation between an azide and a terminal alkyne through Staudinger-Betozzi ligation or alkyne-azide cycloaddition is currently the most reliable and biocompatible reaction to "click" biomolecules with diverse chemical functionalities or fluorescent molecules.²¹⁻²³ The distinct chemical and biochemical advantages of the benzophenones and azides make them attractive to be incorporated into the protein polymer backbone as photo-initiated cross-linkers.²⁴⁻²⁸

In this study, we are particularly interested in the incorporation of the photo-crosslinkers that derived from phenylalanine analogues, *para*-benzoyl-L-phenylalanine (Bpa)^{17,29}, and *para*-azido-L-phenylalanine (AzF)³⁰. Upon the excitation in the near UV wavelength, AzF and Bpa can generate radical species and subsequently react with the protein side chains in the close vicinity (Fig. 1).³¹ Although these photo-initiated probes have been widely used to crosslink small molecules or peptides with proteins, there are no simple methods to assist the incorporation of these probes at multiple defined sites within the larger proteins as well as protein polymers, which are inaccessible by solid-phase synthesis. Recently, Schultz and coworkers have developed the selections of mutant *Methanococcus jannaschii* TyrRS (MjTyrRS) that aminoacylates the mutant RNA^{Tyr}_{CUA} with a set of unnatural amino acids.³² In their later publications, they have demonstrated the selection of orthogonal aaRSs that specifically incorporate the Bpa (BpaRS)²⁹ and AzF analogues (AzFRS)³⁰ into proteins in response to the amber codon UAG in *E. coli* system. They performed *in vitro* and *in vivo* photo-crosslinking experiments on the model protein, glutathione *S*-transferase (GST), which was converted

into a covalently linked homo-dimer initiated by the analogues located specifically at the dimer interface. This innovative development allows the incorporation of a site-specific photo-crosslinker into virtually any protein that can be expressed in *E. coli*, thereby promoting the *in vivo* or *in vitro* crosslinking of proteins. However, the development of practical labeling strategies using this "expanded genetic code" approach depends on the efficiency and fidelity of analogue incorporation. Although amino acid analogues are now typically incorporated into proteins with acceptable efficiency and excellent fidelity by the amber suppression strategy, further system optimization with increased protein

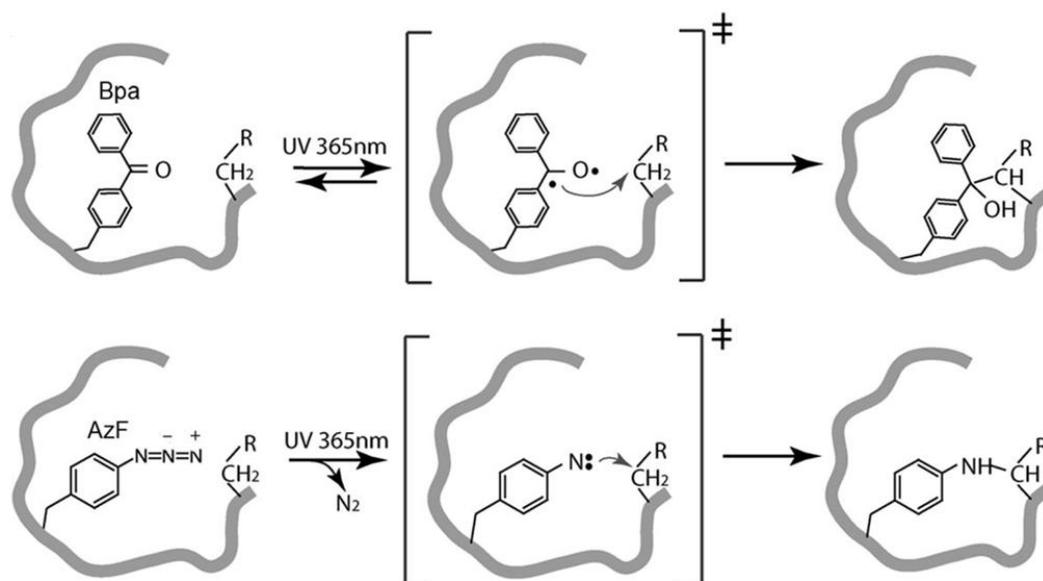


Figure 1. Bpa and AzF serve as photo-crosslinkers. When exposed to the UV light, Bpa and AzF generate bi-radicals (intermediate states) and form covalent linkages with nearby side-chains.³¹

yields is highly desirable, especially for multi-site substitution.

Improvements in the components of the expression system have been demonstrated in the optimization of the orthogonal aaRS/tRNA expression vector. Young et al.³³ have constructed a series of plasmids and showed that an increased expression of

orthogonal aaRS/tRNA pair was effective in decreasing protein truncation and increasing overall yields of mutant proteins. A plasmid with mid-copy-number to low-copy-number p15A origin was finally chosen to encode the orthogonal aaRS/tRNA pair. Farrell and coworkers have demonstrated the improved yields of Bpa-containing polypeptides via the development of the plasmid, pDULE-Bpa plasmid, co-expressing the *M. jannaschii* amber suppressor tRNA and mutant synthetase.³⁴ In a later publication, Stokes et. al. further engineered an evolved Bpa-RS with enhanced activity and also an UAA-RS with broad substrate specificity toward a set of Bpa-related derivatives.³⁵ On the other hand, another important aspect to optimize the orthogonal system is to enhance the affinity of the orthogonal tRNA toward the orthogonal synthetase or other translational machinery. Wang and coworkers have demonstrated that optimizing the affinity between anti-codon recognition region of the orthogonal tRNA and synthetase significantly increased the incorporation efficiencies of various unnatural amino acids.³⁶ Furthermore, since the mutRNA^{Tyr}_{CUA} is derived from an archaeal tRNA that is significantly different in sequence from *E. coli* tRNAs, it may not function optimally with the *E. coli* translational machinery. During protein synthesis, each tRNA must be selectively charged by its cognate aaRS and the resulting aminoacyl-tRNA has to be efficiently bound to the elongation factor Tu (EF-Tu) for the transport to the ribosome. Misacylation of tRNA has recently been shown to perturb the binding to the ribosomal A site.³⁷ While correctly acylated-tRNAs have nearly uniform binding affinity for EF-Tu, tRNAs bearing non-cognate amino acids show a broad range of affinities³⁸, which indicate that both the tRNA and the esterified amino acid can influence EF-Tu binding. The structure of the non-natural amino acids can affect the ability of the corresponding aminoacyl-tRNA to

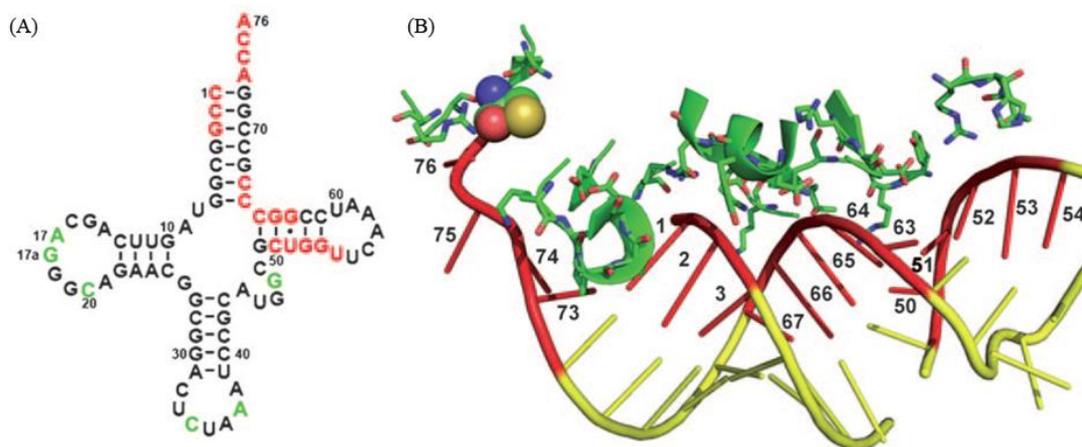


Figure 2. The EF-Tu/ tRNA interface. (a) Putative *M. jannaschii* tRNA^{Tyr}_{CUA} nucleotides that interact with EF-Tu are shown in red. Previously mutated positions are shown in green. (b) A diagram derived from the X-ray crystal structure of *T. aquaticus* EF-Tu/*E. coli* cysteinyl-tRNA^{Cys}. The tRNA residues that interact with EF-Tu are numbered and showed in red.⁴⁰

function productively during translation. Recently, it was also shown that the mutation of the EF-Tu amino acid binding pocket resulted in a more efficient binding of tRNAs bearing bulky nonnatural amino acids.³⁹ Thus, it is likely that tRNAs acylated with non-cognate amino acids result in lower efficiency of protein synthesis because of their loose interactions with EF-Tu or the ribosome. Schultz's group has identified specific mutations in *M. jannaschii* suppressor tRNA to afford an increased expression of mutant proteins in *E. coli*.⁴⁰ Directed evolution experiments focused on the T-stem of MjtRNA^{Tyr}_{CUA} (Fig. 2) identified several modified suppressors with increased unnatural amino acid incorporation efficiently toward several MjaaRS. While most identified tRNAs show some improvement in the protein yields for all aaRS tested, the degree of the improvement for each tRNA often varies depending on the amino acid analogues.

However, these evolved tRNAs could facilitate the creation of an optimized system for the genetic incorporation of the photo-crosslinkers in *E.coli* system.

In this chapter, we optimize the amber suppression strategy for the incorporation of photo-crosslinkers, Bpa and AzF, by utilizing evolved amber suppressor tRNA/ aaRS pairs. We have tested two evolved tRNAs in combination with engineered synthetases, BpaRS and AzFRS, derived from the MjTyrRS. We hypothesize that the optimized orthogonal tRNA/synthetase pair could afford higher activity to function in the RF1-attenuated *E. coli* expression strain MRA30⁴¹ and incorporate the selected photo-crosslinker at multiple amber positions within the growing polypeptide chain. This new system that utilizing *E. coli* MRA30 host with the optimized suppressor tRNAs affords higher yields of mutant protein synthesis. In this system, new photo-crosslink experiments with multiple photo-probes could be designed to investigate the details of protein interaction networks. In addition, the enhanced incorporation should permit high-level biosynthetic incorporation of the photo-crosslinkers with high yields of elastin-mimetic polypeptides. Therefore, a variety of protein-based materials can be genetically designed and synthesized with retained mechanical bioactivities and extended photo-crosslinkable functionalities.

Experimental Methods

Materials

Chemical reagents were purchased from Fisher Scientific, Inc. (Pittsburgh, PA) or Sigma Chemical Co. (St. Louis, MO) unless otherwise noted. The amino acid derivatives *para*-benzoyl-*L*-phenylalanine (Bpa) and *para*-azido-*L*-phenylalanine (AzF) were purchased Bachem Bioscience, Inc. (King of Prussia, PA). Isopropyl- β -D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, and deoxynucleotide solution mix (dNTPs) were purchased from New England Biolabs, Inc. (Beverly, MA) and *Pfx* DNA polymerase was obtained from Invitrogen Corp. (Carlsbad, CA). The pSU81 plasmid was a gift from Professor William McClain at the University of Wisconsin.⁴² Plasmid pHEC2 was synthesized by Dr. Holly Carpenter in the Conticello lab at Emory University.⁴³ Plasmid pDULE-BpaRS was generously provided by Dr. Ryan Mehl of Franklin and Marshall College (Lancaster, PA).³⁵ Single-stranded, PAGE-purified oligonucleotides were obtained from either Sigma-Genosys, Inc. (The Woodlands, TX) or Integrated DNA Technologies (Coralville, IA). Total synthesis of codon-optimized genes encoding wild-type *M. jannaschii* tyrosyl-tRNA synthetase (MjTyrRS) and *M. jannaschii* azidophenylanyl-tRNA synthetase (MjAzFRS) were performed by DNA2.0, Inc. (Menlo Park, CA). The *E. coli* strains, TOP10F' and DH10B, were obtained from Invitrogen Corp. (Carlsbad, CA) and the *E. coli* strain MRA30 was kindly provided by permission of Dr. Monica Rydén-Aulin with the assistance of Jaroslav Belotserkovsky.⁴¹ Plasmids and primers used in this chapter are listed in Table 1 and Table 2 respectively.

General Methods

Basic molecular biology techniques, including polymerase chain reaction, ligation, immunoblotting, and gel electrophoresis, were adapted from a standard molecular cloning manual 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 (0.22 μ m 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. Site-directed mutagenesis was performed using Stratagene's (La Jolla, CA) Quick-Change mutagenesis technique from gene-specific oligonucleotide primers. *E. coli* strains were grown in Luria-Bertani (LB) medium with appropriate antibiotics at 37 °C, 200 rpm otherwise noted. Plasmid DNA preparation and purification were performed using the QIAprep Spin Miniprep Kit (QIAGEN Inc., Valencia, CA), and the DNA Clean & ConcentratorTM (Zymogen, CA). Polymerase chain reaction (PCR) was carried out using a MJ MiniTM Gradient Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules, CA). The PCR products or plasmids were digested with restriction enzymes and separated by 1-4% agarose TBE gels. The desired DNA fragments were purified with ZymocleanTM Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA). Ligations were carried out in a 10 μ L volume with 0.5 μ L T4 DNA Ligase (200 units) and 1X T4DNA Ligase buffer at 16 °C overnight. Chemically competent cells of TOP10F' or DH10B were transformed with ligation mixtures, recovered at 37 °C for an hour. The recovery mixture (200 μ L) was spread onto LB agar media supplemented with appropriate antibiotics. The plates were incubated at 37 °C for

12 to 14 hours. Proteins were purified by metal-affinity chromatography using the cobalt-charged TALON[®] Metal affinity Resin from Clontech Laboratories (Mountainview, CA).

Construction of the *M. jannaschii* tRNA plasmid

A duplex DNA cassette encoding optimized MjtRNA_{CUA} (Opt)³³ or MjtRNA_{CUA} (Nap1)⁴⁰ was synthesized by annealing the synthetic single-strand oligonucleotide cassettes. The DNA cassette was phosphorylated by T4 polynucleotide kinase at 37 °C for 1 h and purified by Zymogen DNA Clean & Concentrator[™] kit. The synthetic tRNA was cloned into the compatible *EcoR* I/ *Pst* I sites within the plasmid pSU81. The pSU81 plasmid, containing the MjtRNA_{CUA} (Opt) flanked by the *lpp* tRNA promoter and *rrnC* tRNA terminator sequences was sequenced by using pSU81-F/ R primers and assigned as pIL83. The pSU81 plasmid that contains the MjtRNA_{CUA} (Nap1) was verified and designated as pIL84. The synthetic DNA encoding the tRNA expression cassette from pIL83 and pIL84 was double digested with *Nhe* I and *Pvu* I, and cloned into the compatible sites within pHEC2 to afford plasmid pIL86 and pIL87. The MjtRNA_{CUA} gene was confirmed by automated DNA sequencing analysis using HtRNA-F/ R primers and under the transcriptional control of the *lpp* promoter and *rrnC* terminator.

Construction of the *M. jannaschii* aminoacyl-tRNA synthetase / tRNA plasmid

The synthetic gene encoding the wild-type MjTyrRS was optimized for the expression in *E. coli* and synthesized by DNA2.0 (Menlo Park, CA) to produce pJ201-MjTyrRS. The MjTyrRS gene was released by double digestion with *Kpn* I and *Xba* I, and cloned into the acceptor plasmid pIL86 or pIL87 that already contained the MjtRNA_{CUA} expression cassette. Synthetic DNA sequences were verified by automated

DNA sequencing analysis using 5-Pro/ 3-Pro primers and the plasmids were assigned as pIL88 and pIL89. The wild-type MjTyrRS gene in the final expression plasmid is under the control of the constitutive $P_L tet$ promoter.

The optimized BpaRS with the following mutation sites: Y32G, E107S, D158T, I159S, was introduced into the MjTyrRS gene template by double-PCR mutagenesis from gene-specific primers. The PCR products were digested with *Kpn* I and *Xba* I and cloned into the MjtRNA_{CUA} plasmids pIL86 or pIL87 to generate plasmid pIL90 and pIL91. The advanced BpaRS with the following mutation sites: A31V, Y32G, E107P, D158S, I159S, was introduced into the MjTyrRS gene template using double-PCR mutagenesis from gene-specific primers. The PCR products were digested with *Kpn* I and *Xba* I, and cloned into the MjtRNA plasmid pIL86 or pIL87 to afford plasmids pIL96 and pIL97. The recombinant plasmids of interest were selected and sequenced by 5-Pro/ 3-Pro primers to confirm the location of mutations.

The synthetic gene encoding the MjAzFRS with introduced mutations (Y32T, E107N, D158P, I159L, L162Q) based on wild-type MjTyrRS, was optimized for the expression in *E. coli* and synthesized by DNA2.0 (Menlo Park, CA) to produce pJ201-AzFRS. The MjAzFRS gene was excised by *Kpn* I and *Xba* I digestion, and cloned into compatible sites within plasmids pIL86 and pIL87 that already had the MjtRNA_{CUA} gene expression cassette to afford plasmid pIL100 and pIL101. The sequences were confirmed by automated DNA sequencing analysis using 5-Pro/ 3-Pro primers.

Bacterial growth and expression

Electro-competent cells of *E. coli* strain MRA30 were prepared. Cells were transformed with the plasmids, pIL80 (sfGFP-(UAG)₃), pIL43.5 (elastin-(UAG)₁₂or

pIL43.11 (elastin-(UAG)₂₂), encoding the target protein and selected on LB agar supplemented with ampicillin (100 µg/mL) via incubation at 30 °C for 12-14 h. Fresh electro-competent cells with individual target protein were prepared from single colonies for further transformation. The plasmid DNA containing the MjTyrRS pair, MjBpaRS pair, MjAzFRS pair, or the empty plasmid, pHEC2 (lacking the synthetase/tRNA pair), was then transformed and plated on LB agar media supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) to generate expression strains for these studies. The plasmid pDULE-Bpa was also transformed into the pIL80-MRA30 strain and plated on LB agar supplemented with ampicillin (100 µg/mL) and tetracycline (25 µg/mL) as a control study.

Single colonies of the expression strains were inoculated into sterile LB broth supplemented with the appropriate antibiotics (100 µg/mL ampicillin and 34 µg/mL chloramphenicol or 25 µg/mL tetracycline) as required for plasmid maintenance. One liter of LB or TB supplemented with the appropriate antibiotics was inoculated with 20 mL of overnight culture and incubated at 30 °C until OD₆₀₀ reached between 0.6 and 0.8 absorbance units. Phenylalanine analogues and IPTG were added to a final concentration of 1 mM to induce the expression of the sfGFP and elastin derivatives. After 12 h, the cells were harvested by centrifugation at 4000 g and 4 °C for 20 min. The cell pellet was re-suspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 8.0) and stored at -80°C.

TALON[®] metal-affinity column purification

The frozen cells were lysed by three freeze/thaw cycles. Lysozyme (1 mg/mL), EDTA-free protease inhibitor cocktail, benzonase (25 units/mL), and MgCl₂ (1 mM)

were added to the lysate and the mixture was incubated shaking at 4 °C for 24 h. The cell lysate was centrifuged at 14,000g for 30 min at 4 °C. For the purification of the soluble sfGFPs and the elastin-mimetic proteins with AzF analogues, the supernatant was loaded directly onto a column containing cobalt-charged TALON[®] resin (2 mL) and washed with lysis buffer (20 mL) containing 20 mM imidazole. The target protein was eluted with elution buffer (20 mL, 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0) and dialyzed (MWCO = 10 kDa) against distilled de-ionized water (5x 4 L). The dialysate of sfGFPs was concentrated by ultrafiltration using Amicon filters (NMWL 10 kDa) to a final volume of 1 mL. Protein concentration was measured by the Bradford assay. The dialysate of elastin-mimetic proteins was lyophilized to produce a white, spongy solid and the yield was measured by dry weight.

The insoluble elastin-mimetic proteins with Bpa analogues were purified from the cell pellet under denaturing conditions. The pellet was suspended in denaturing lysis buffer (50 mL, 50 mM sodium phosphate, 100 mM NaCl, 6 M Urea, pH 8.0) and incubated at 4 °C with shaking overnight. The cell suspension was centrifuged at 14,000 g for 30 min at 4 °C. The cell lysate was then loaded onto the TALON[®] column (3mL) and washed with denaturing lysis buffer (30 mL) containing 20 mM imidazole and 6 M urea. The target protein was eluted with denaturing 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.

Flow cytometry

Aliquots (1 mL) of *E. coli* cells from expression cultures were grown until the OD₆₀₀ reached approximately 1.0 absorbance units. The cultures were centrifuged at 4000 g and 4 °C for 10 min and re-suspended in 1 mL of phosphate-buffered saline pH 7.4. Flow cytometry was performed using a LSRII flow cytometer (Beckton Dickinson) equipped with a 100 mW solid-state laser emitting at 488 nm for the excitation of sfGFP, a 505 nm LP dichroic mirror and a 530/30 bandpass filter. Forward scatter (FSC), sideward scatter (SSC) and green fluorescence 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 with different non-canonical amino acids. Data was analyzed using FlowJo software (TreeStar.com).

Thermolysin digestion

Protein derivatives were dissolved in sterile water at a concentration of approximately 1 mg/mL. Dithiothreitol (DTT) was added as a final concentration of 10 mM and the mixture was incubated at 100 °C for 30 min to denature the protein. After the reaction mixture was cooled to 37 °C, thermolysin was added to the denatured protein solution to a concentration of 1: 50 (w/w) ratio with respect to protein and the reaction mixture was incubated at 37 °C for 12 h. The products from the proteolysis reaction were passed through a PepClean™ C18 spin column (Thermo Scientific, Inc.) to remove the salts and thermolysin.

Mass spectrometry

Electrospray mass spectra of sfGFP derivatives were acquired by LTQ-FT mass spectrometer (ThermoElectron, San Jose, CA) in the positive ion mode (200-2000 m/z) using a spray voltage of 4 kV, a capillary voltage of 41 V, a sheath gas flow rate of 20 arbitrary units. The tube lens voltage was 215V, and the AGC setting was $5e+05$. The Xtract program in Xcalibur (ThermoElectron, San Jose, CA) was used for spectral deconvolution. MALDI-TOF experiments were performed on an Applied Biosystems® Voyager™ System 428 mass spectrometer (Life Technologies Corporation; Carlsbad, CA) in the linear positive ion mode. Ferulic acid was used as matrix at a concentration of 10 mg/mL in a mixture of 75 % acetonitrile (ACN) and 0.2% formic acid (FA) in deionized water. The elastin-mimetic protein (1 mg/mL in distilled water) was mixed with the matrix solution in a ratio of 1:10. Two-microliters of the mixture were spotted on a stainless steel, 100-position flat sample plate and dried in air. The Sequazyme™ bovine serum albumin (BSA) standards test kit (Applied Biosystems Inc.; Carlsbad, CA) was used as a standard for external calibration with the sinapinic acid matrix solution. The sinapinic acid (3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid) was used at a concentration of 20 mg/mL in a mixture of 50 % ACN and 0.1 % FA in deionized water.

Tandem MS of selected thermolysin-digested peptides were carried out using both ESI and MALDI mass spectrometry. Electrospray measurements were acquired by LTQ-FT mass spectrometer (ThermoElectron, San Jose, CA) in positive ion mode using a spray voltage of 4 kV, a capillary voltage of 44 V, a sheath gas flow rate of 20 arbitrary units, a CID collision energy 25-35eV. The tube lens voltage was 120 V, and the AGC setting was $1e+06$. Solutions of each peptide (1×10^{-5} M, methanol) were infused with a

rate of 5 $\mu\text{L}/\text{min}$. Ultra pure helium was the collision gas in MS^n experiments carried out in the ion trap. Thirty to fifty scans were recorded and averaged for accurate mass measurements. MALDI spectra were recorded on an Ultraflex TOF/TOF instrument (Bruker Daltonics GmbH, Bremen, Germany) equipped with LIFT capability. A pulsed Nd:YAG laser at a wavelength of 355 nm was operated at a frequency of 100 Hz. The source was operated in the positive mode with an acceleration voltage of 25.0 kV, and a delayed extraction time of 20 ns was applied. A solution of the alpha-cyano-4-hydroxycinnamic acid (4-CHCA) matrix at a concentration of 10 mg/ml in 50 % ACN was mixed with the peptide sample in equal amount. Mass spectra were acquired from 100 laser shots in the range m/z 500 to 3000 in the reflectron mode. MS/MS experiments were performed under laser-induced dissociations (LID) conditions with the LIFT cell voltage parameters set at 19.0 kV for a final acceleration of 29.5 kV (reflector voltage). Data analysis of the MS/MS data was carried out using the package Bio Tools™ (Bruker Daltonics, Bremen, Germany).

Photo-crosslink experiments

Model photo-crosslinking experiments were carried out on selected AzF-derivative proteins at 4 °C or 37 °C. SfGFP-(AzF)₃ and elastin-(AzF)₁₂ were dissolved in sterile water at a concentration of approximately 1 mg/mL. Ultraviolet (UV) irradiation was performed using a hand-held UV lamp at a wavelength of 356nm for 30 min. Cross-linked and non-crosslinked proteins were assayed via the migration in gel electrophoresis.

Table 1. Plasmids and strains utilized in Chapter 3.

Plasmids	Relevant characteristics	Reference
pIL7	sfGFP-CCC in pIL2, Amp ^R	Chapter 2
pIL80	sfGFP-UAG in pIL2, Amp ^R	Chapter 2
pJ201-MjTyrRS	MjTyrRS in pJ201 (DNA2.0), Kan ^R	This study
pJ201-MjAzFRS	MjAzFRS in pJ201 (DNA2.0), Kan ^R	This study
pIL43.5	Elastin-UAG _{12mer} concatemer in pIL5, Amp ^R	Chapter 2
pIL43.11	Elastin-UAG _{22mer} concatemer in pIL5, Amp ^R	Chapter 2
pSU81	Synthetic plasmid containing a tRNA expression cassette, Cm ^R	Reference ⁴²
pHEC2	Modified pPROTet.E133/LarA.231, Cm ^R	Reference ⁴³
pIL83	MjtRNA Opt in pSU81, Cm ^R	This study
pIL84	MjtRNA Nap1 in pSU81, Cm ^R	This study
pIL86	MjtRNA Opt expression cassette in pHEC2, Cm ^R	This study
pIL87	MjtRNA Nap1 expression cassette in pHEC2, Cm ^R	This study
pIL88	MjTyrRS in pIL86 (MjtRNA Opt), Cm ^R	This study
pIL89	MjTyrRS in pIL87 (MjtRNA Nap1), Cm ^R	This study
pIL90	MjBpaRS in pIL86 (MjtRNA Opt), Cm ^R	This study
pIL91	MjBpaRS in pIL87 (MjtRNA Nap1), Cm ^R	This study
pIL96	Advanced BpaRS in pIL86 (MjtRNA Opt), Cm ^R	This study
pIL97	Advanced BpaRS in pIL87 (MjtRNA Nap1), Cm ^R	This study
pDULE-Bpa	Advanced BpaRS with MjtRNA _{CUA} , Tet ^R	Reference ³⁵
pIL100	MjAzFRS in pIL86 (MjtRNA Opt), Cm ^R	This study

pIL101	MjAzFRS in pIL87 (MjtRNANap1), Cm ^R	This study
Strain	Genotype	Reference
Top10F'	F [lacIq, Tn10(TetR)] <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) Φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139</i> Δ(<i>ara leu</i>) 7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i>	Invitrogen
DH10B	F <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 endA1 araD139</i> Δ (<i>ara, leu</i>)7697 <i>galU galK λ- rpsL nupG</i> /pMON14272 / pMON7124	Invitrogen
MRA30	<i>prfA1 recA rph</i>	Reference ⁴¹

Table 2. Primers utilized in Chapter 3.

Name	Sequence (5'→3')
MjtRNAOpt-F	AATCCC GGCGGTAGTTCAGCAGGGCAGAACGGCG GACTCTAAATCCGCATGGCAGGGGTTCAAATCCCC TCGCGCGGACCACTGCA
MjtRNAOpt-R	GTGGTCCGGCGGAGGGGATTTGAACCCCTGCCATG CGGATTTAGAGTCCGCCGTTCTGCCCTGCTGAACT ACCGCCGGG
MjtRNANap1-F	AATCCC GGCCGTAGTTCAGCAGGGCAGAACGGCG GACTCTAAATCCGCATGGCATGGGTTCAAATCCCA TCGCGCGGACCACTGCA
MjtRNANap1-R	GTGGTCCGGCCGATGGGATTTGAACCCATGCCATG CGGATTTAGAGTCCGCCGTTCTGCCCTGCTGAACT ACGGCCGGG
BpaRSY32G-F	GAAGAGCGCGGGCATTGGCTTCGAGCC
BpaRSY32G-R	GGCTCGAAGCCAATGCCCGCGCTCTTC
BpaRSE107S-F	GTATGGCTCTAGCTTTCAGCTGGAC
BpaRSE107S-R	GTCCAGCTGAAAGCTAGAGCCATAC
BpaRSD158TI159S-F	GCAAGTTAACACCAGCCACTACCTGGGCG
BpaRSD158TI159S-R	CGCCCAGGTAGTGGCTGGTGTTAACTTGC
BpaRSE107P-F	GTATGGCTCTCCGTTTCAGCTGGAC
BpaRSE107P-R	GTCCAGCTGAAACGGAGAGCCATAC
BpaRSD158TSI159S-F	GCAAGTTAACAGCAGCCACTACCTGGGCG
BpaRSD158TSI159S-R	CGCCCAGGTAGTGGCTGCTGTTAACTTGC
BpaRSA31VY32G-F	GATGAGAAGAGCGTGGGCATTGGCTTCGAGCC
BpaRSA31VY32G-R	GGCTCGAAGCCAATGCCACGCTCTTCTCATC

Results and Discussion

Incorporation of the photo-crosslinkable amino acids into protein-based materials is limited for either the length of peptides for solid-phase synthesis or the suppression efficiency of current amino-acyl synthetase (aaRS) pairs. In this chapter, we optimize the orthogonal aaRS pairs with enhanced tRNA suppressors to achieve higher incorporation efficiency for the amino acid analogues, *para*-benzoyl-*L*-phenylalanine (Bpa), and *para*-azido-*L*-phenylalanine (AzF). In order to monitor the permissive efficiency of the optimized aminoacyl-RS/tRNA pairs, a triple non-sense insertion variant of green fluorescent protein, described in Chapter 2, was adapted as a reproducible reporter for full-length protein expression. Fluorescence-activated cell sorting (FACS) was employed to monitor the incorporation efficiency of the orthogonal pairs toward the analogues.

Previous studies in our lab have demonstrated the incorporation of Bpa analogues utilizing the pDULE-BpaRS plasmid.⁴⁴ However, stable transformants were not obtained in *E. coli* strain MRA30 and the protein expression was unstable and insufficient. In this chapter, we consolidated the newly optimized orthogonal pair with an evolved tRNA into a single highly adaptable vector system⁴³ under the transcription control of the *P_{Ltet}* promoter. The genetic modified RF1-attenuated *E. coli* MRA30 strain, which has been characterized by the increased read-through of the amber codons in Chapter 2, was employed to facilitate the expression of the artificial proteins, sfGFP-(UAG)₃ and elastin-mimetic polypeptides.

Suppression Efficiency Assay

Previous results by Schultz's group have indicated that even with the same evolved tRNA sequence, the efficiency of incorporation depends not only on the local

protein structure but also on the variant of unnatural amino acids.⁴⁰ Therefore, we selected two evolved MjTyr_{CUA} tRNAs, Nap1 tRNA and Opt tRNA, for these studies (Fig. 3). In addition, two engineered Bpa synthetase variants derived from MjTyrRS were engineered and the orthogonal pairs were cloned into the pHEC2 expression vector. As the suppression efficiency also depends on the local sequence context⁴⁵, it is not easy to directly compare the expression experiments from different sites of different proteins. A green fluorescent protein (sfGFP) with amber codon insertions was constructed as a robust reporter system to monitor the difference of the activity from the orthogonal pairs. In this assay, reporter gene and the engineered orthogonal pair were transformed into the *E. coli* MRA30 strain. Protein expression was performed in the presence or absence of the selected amino acid analogues to demonstrate the fidelity of the incorporation with various aaRS pairs. Suppression of the UAGs by the orthogonal pair results in the full-length fluorescent sfGFP in which the total fluorescence intensity of cells measured by fluorescence-activated cell sorting (FACS) reports the incorporation efficiency of the synthetase with the evolved tRNA.

We constructed four orthogonal BpaRS pairs for the optimization (Fig. 4). Compared to the parent pDULE-BpaRS, the fluorescence intensity of the Nap-1 tRNA pairs (pIL96 and pIL97) was enhanced to around 230~300%, while the Opt tRNA pairs (pIL90 and pIL91) increased to 200~230% (Fig. 5). In addition, a control expression was performed using a conventional *E. coli* host strain DH10B with unaltered release factor recognition and the fluorescence response of the sfGFP-(UAG)₃ in the DH10B strain was indistinguishable from the background. The presence of the tyrosine in the LB medium, as the native substrate for the wild-type synthetase, did not support a significant level of

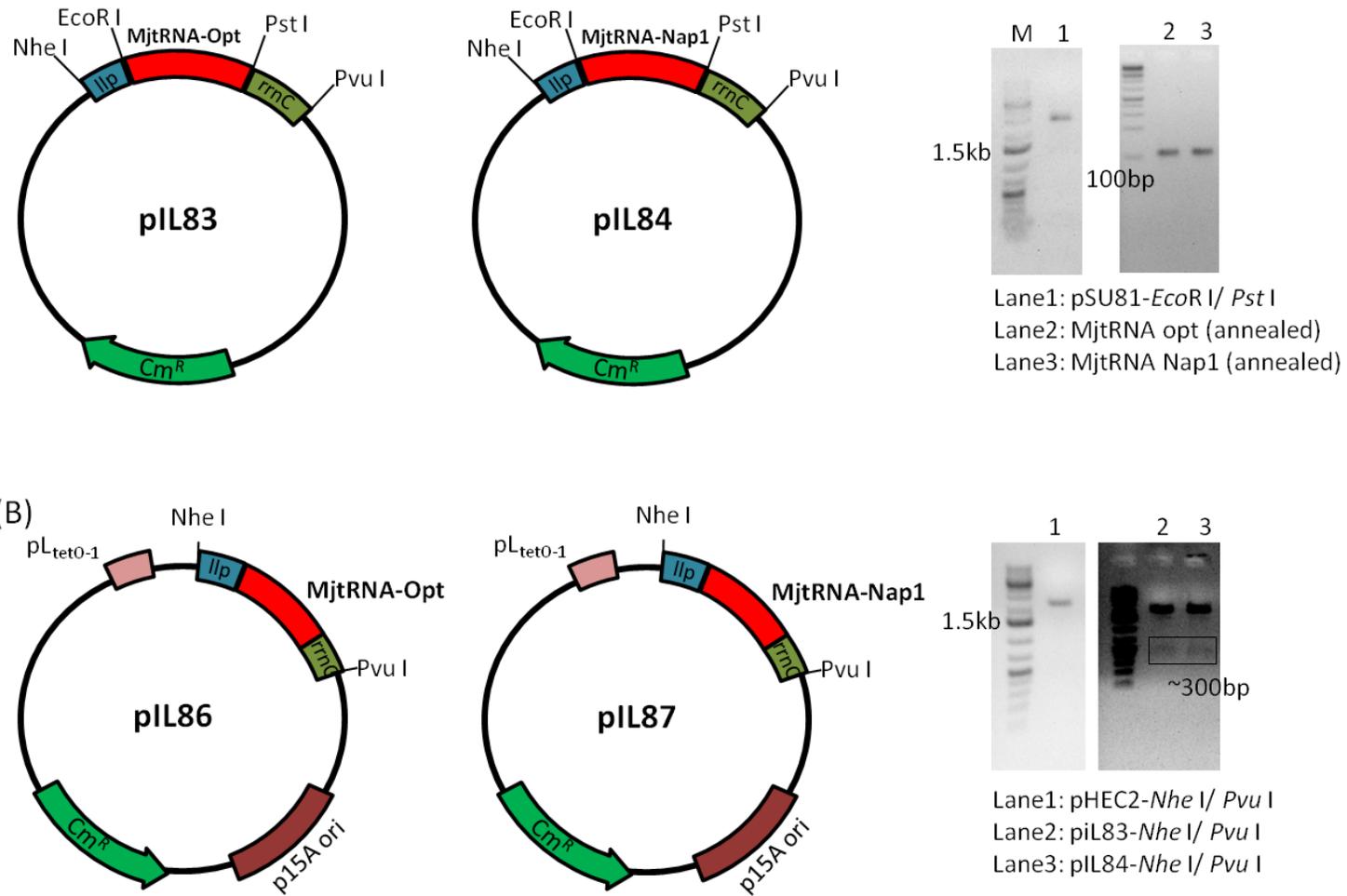


Figure 3. (A) The pIL83 and pIL84 are derived from the pSU81 plasmid and harbor the MjtRNA genes. (B) The pIL86 and pIL87 are derived from the pHEC2 plasmid with the MjtRNA gene expression cassette.

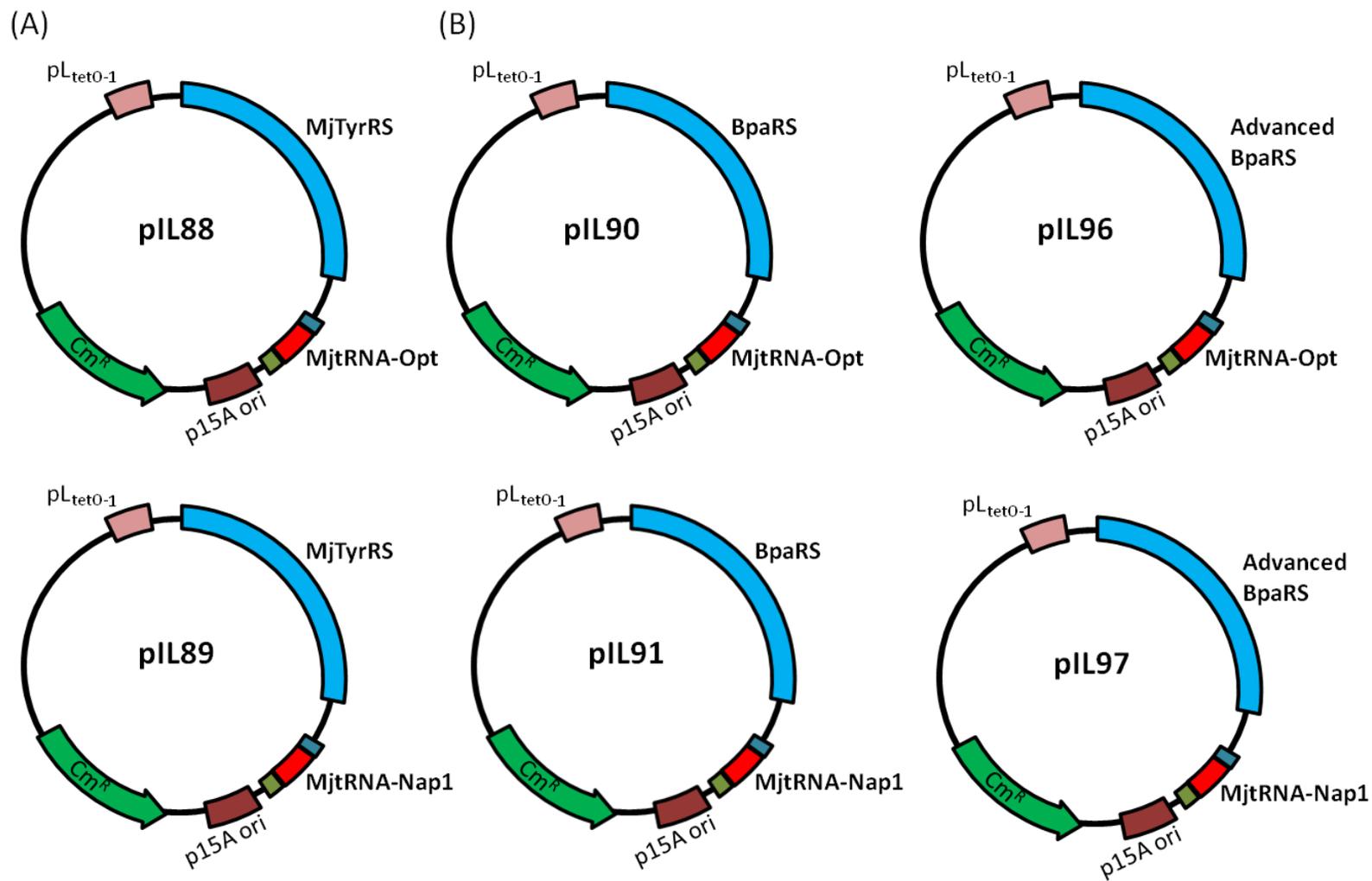


Figure 4. Expression plasmids of MjTyrRS (A) and MjBpaRS (B) with evolved MjtRNA_{CUA} cassette.

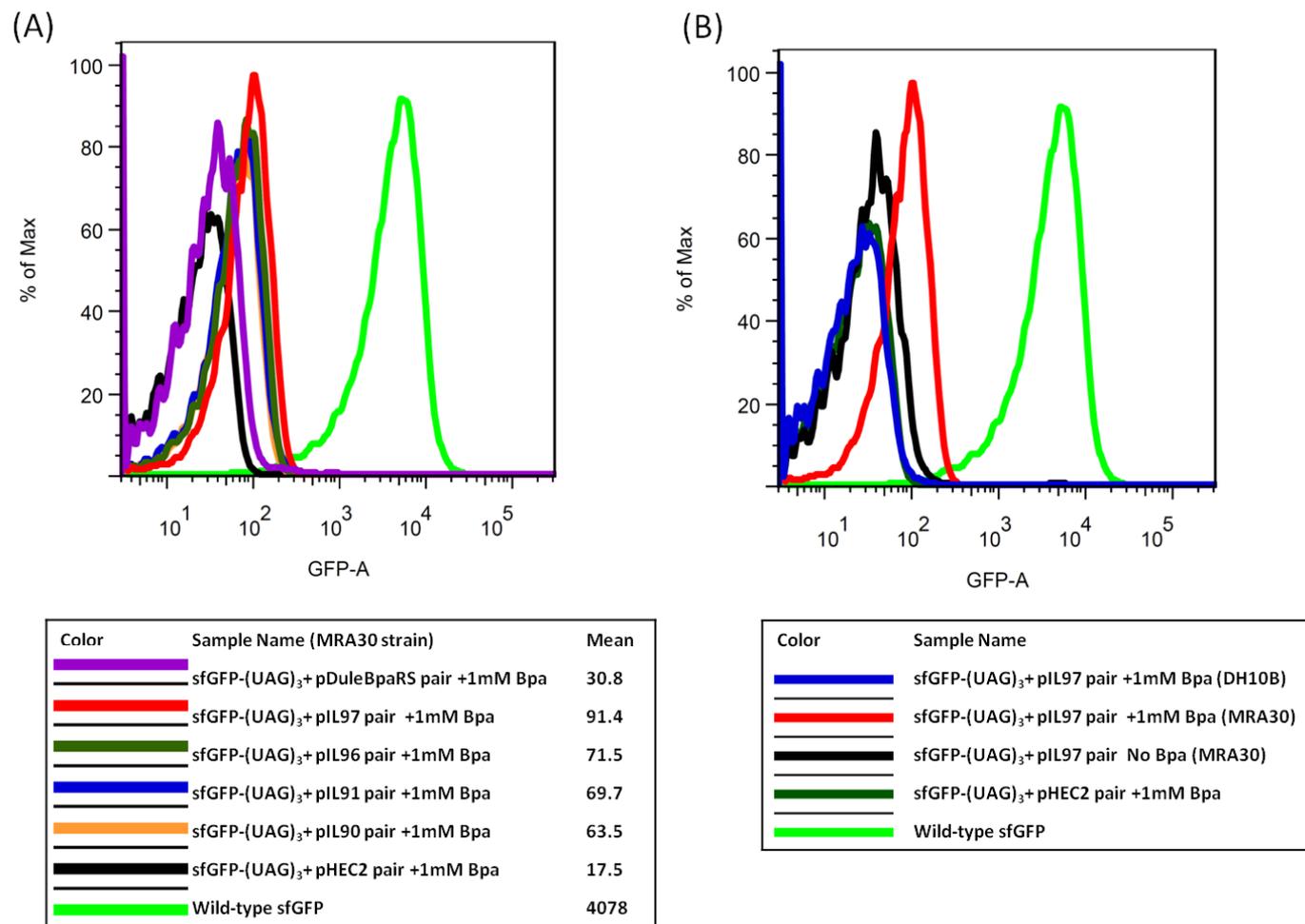
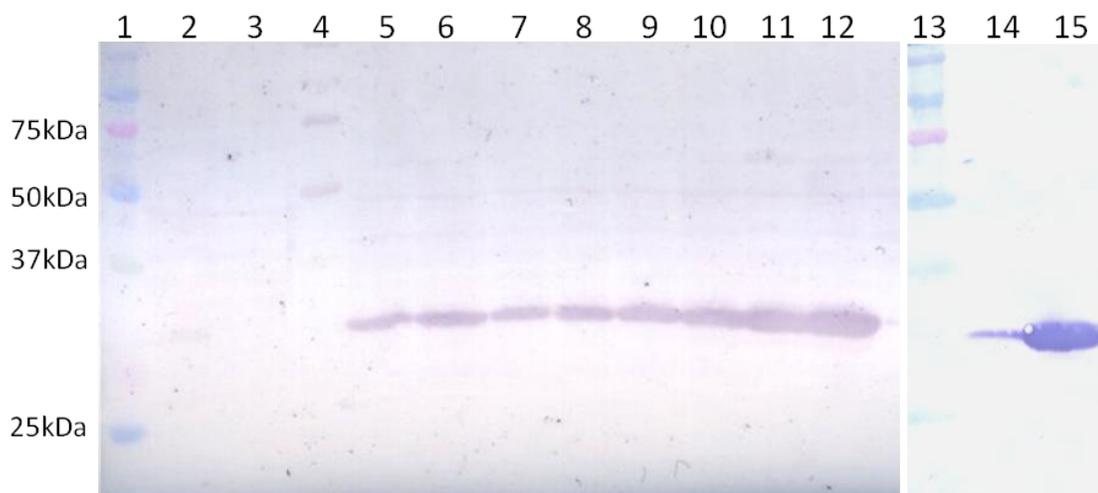


Figure 5. FACS analysis of sfGFP-(Bpa)₃ for the indicated BpaRS pair. (A) Expression of orthogonal BpaRS pairs with Bpa analogues in *E. coli* MRA30 strain and the mean fluorescence. (B) Expression of the orthogonal BpaRS pair, pIL97, in *E. coli* MRA30 and DH10B strain.



Lane1: Precision Plus Dual Color Prestained Standards (Bio-Rad)
 Lane2: MRA30 sfGFP-(UAG)₃+pIL96 without analogues
 Lane3: MRA30 sfGFP-(UAG)₃+pIL97 without analogues
 Lane4: Perfect Protein marker (Novagen)
 Lane5-6: MRA30 sfGFP-(UAG)₃+pIL90 Bpa
 Lane7-8: MRA30 sfGFP-(UAG)₃+pIL91 Bpa
 Lane9-10: MRA30 sfGFP-(UAG)₃+pIL96 Bpa
 Lane11-12: MRA30 sfGFP-(UAG)₃+pIL97 Bpa
 Lane13: Precision Plus Dual Color Prestained Standards (Bio-Rad)
 Lane14: DH10B sfGFP-(UAG)₃+BpaRS (pIL97) Bpa
 Lane15: MRA30 sfGFP-(UAG)₃+BPaRS (pIL97) Bpa

Figure 6. Western blot analysis for the expression of sfGFP-(UAG)₃ with BpaRS pairs in *E.coli* MRA30 and DH10B strains.

sfGFP expression in the presence of the BpaRS/tRNA pair based on the FACS results.

Electrophoresis and western blot analysis of whole-cell lysates derived from the expression cultures also indicated the selective incorporation only in the presence of Bpa analogues. This results supports that the evolved 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. Our data suggests that the advanced BpaRS with Nap1 tRNA pair (pIL97) showed the best activity for Bpa analogues with respect to the negative control which was deficient with the analogue (Fig. 6).

Aminoacyl-tRNA synthetase specific for AzF (AzFRS) incorporation was also engineered into the plasmids with either Opt or Nap1 tRNA (Fig 7). FACS analysis of the expression cultures indicated a significant increase in fluorescence activity for the sfGFP reporter in the presence of AzF in the MRA30 strain versus the negative control (Fig. 8). Based on fluorescence intensity expressed in the LB medium, the AzF-Nap1 pair (pIL101) demonstrated the highest efficiency for AzF incorporation. It also led to the significant suppression of the sfGFP amber mutant even in the conventional DH10B lab strain. Overall, we demonstrated that the Nap1 tRNA could be genetically encoded, processed, and used to read through amber stop codons for the efficient incorporation of both Bpa and AzF analogues. We have confirmed that both of the synthetase pairs are orthogonal in *E. coli* in response to the amber suppression. Based on these results, the optimal orthogonal pairs for Bpa (pIL97) and AzF (pIL101) are selected and utilized in this study for further quantitative and qualitative characterization.

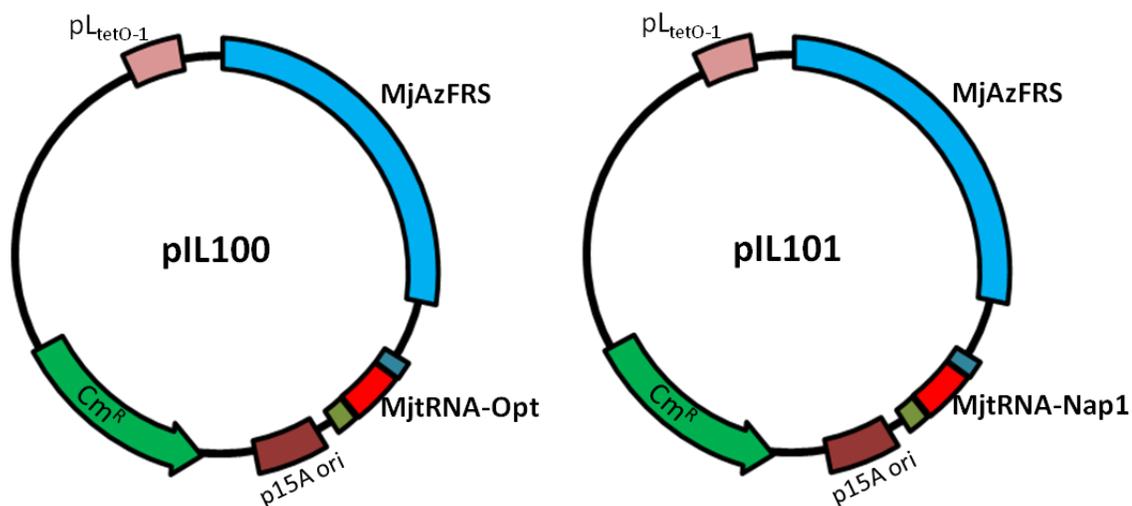


Figure 7. Expression plasmids of MjAzFRS with evolved MjtRNA_{CUA}.

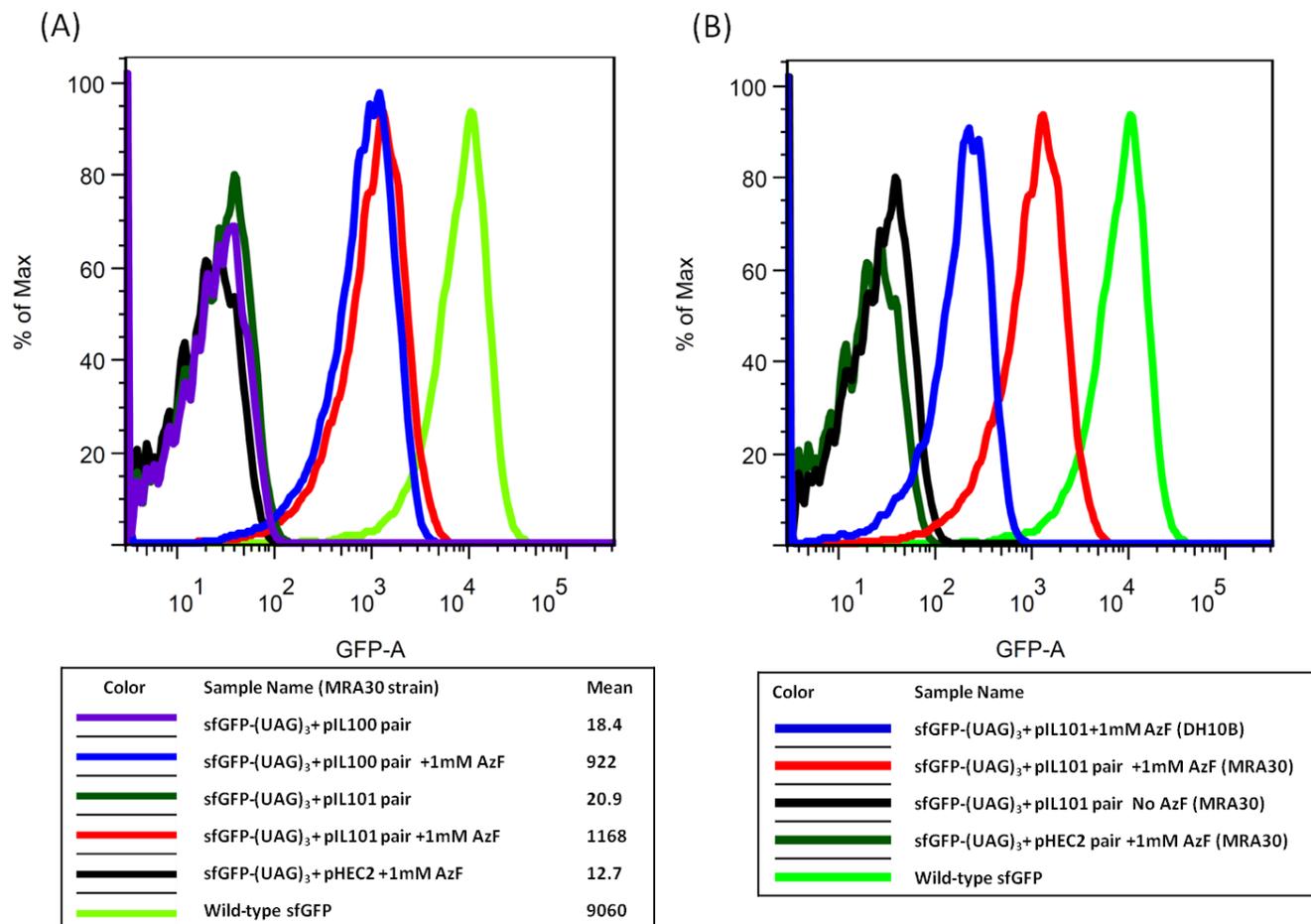


Figure 8. FACS analysis of sfGFP-AzF for the indicated AzFRS pair. (A) Expression of orthogonal AzFRS pairs in *E. coli* MRA30 strain and the mean fluorescence. (B) Expression of orthogonal AzFRS pair, pIL101, in *E. coli* MRA30 and DH10B strains.

The expression level of the sfGFP-(UAG)₃ was assessed under conditions of the co-expression of the BpaRS/Nap1tRNA (pIL96) pair or AzFRS/Nap1tRNA (pIL101) pair as described previously. Preparative-scale expression of sfGFPs was performed in *E. coli* MRA30 in the presence of the respective analogues. The sfGFP derivatives were purified by immobilized metal-affinity chromatography (Fig. 9). The isolated yield of the sfGFP with Bpa analogues was around 4 mg/L. The yield was significantly improved compared to those of the previously developed pDULE-BpaRS system in our lab, which was less than 1 mg/mL. The reduced yield of the sfGFP-(Bpa)₃ compared to the wild-type sfGFP (~60 mg/mL) could be because the target protein remains in the insoluble fraction and binds poorly to the resin (Data not shown). The yield of the sfGFP-(AzF)₃ (13.4 mg/mL) was higher than that of the corresponding sfGFP-(Bpa)₃. The incorporation efficiency includes the differences in the enzymatic activity of the respective aaRS mutants for the non-canonical amino acids, and the suppression efficiency of the charged-suppressor tRNAs during ribosomal chain elongation. Furthermore, deconvoluted ESI-mass spectra of the purified sfGFP derivatives indicated a mass difference between the variant sfGFP and the wild-type sfGFP that corresponded to the mass of three Bpa or AzF residues addition within the polypeptide backbone (Fig. 10 and Table 3). These data suggest that this optimized system that has adapted the evolved Nap1 tRNA and *E. coli* MRA30 strain could benefit future applications by allowing greater reproducibility for the multi-site incorporation of these photo-crosslinkers.

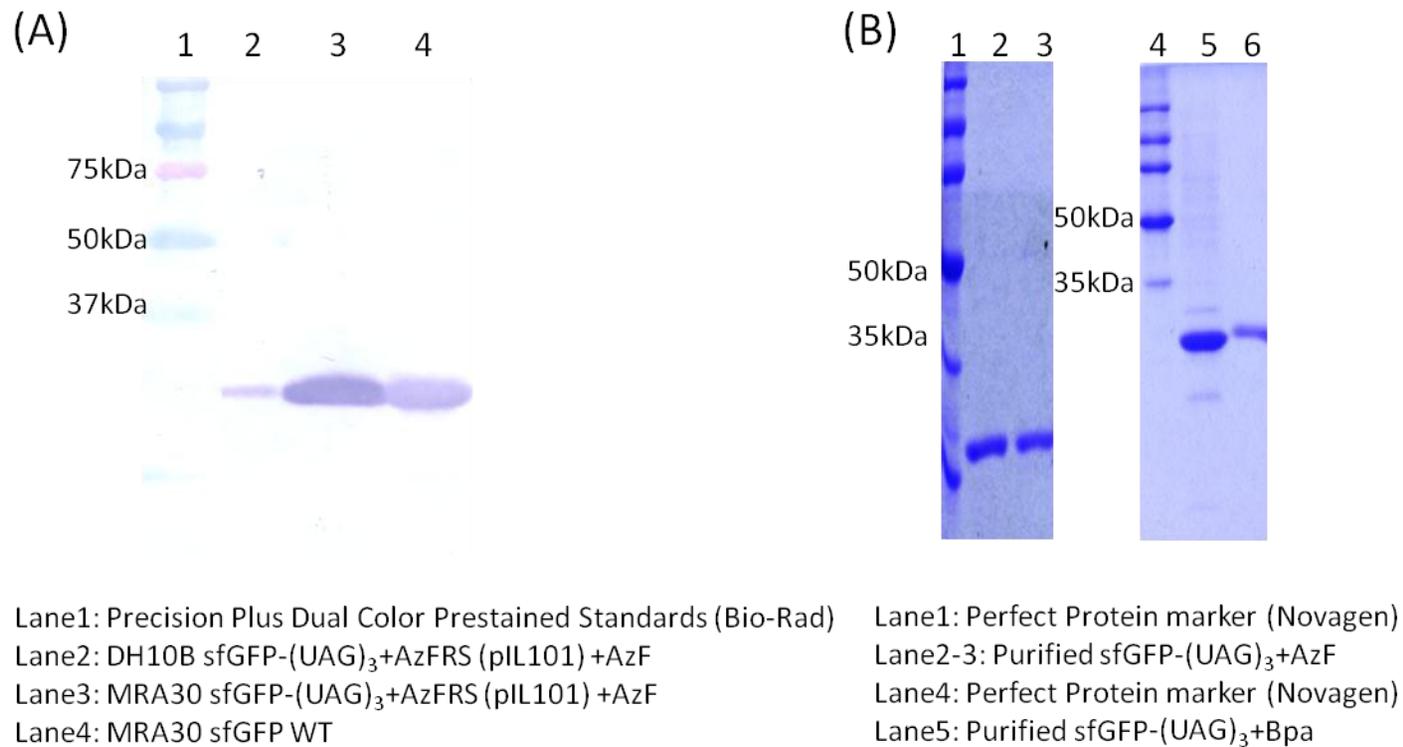


Figure 9. (A) Western blot analysis of the expression of sfGFP-(UAG)₃ with AzFRS pair in the presence of AzF analogues in *E.coli* MRA30 and DH10B strains. (B) SDS-PAGE analysis of the isolated sfGFP-(AzF)₃ and sfGFP-(Bpa)₃ from *E.coli* MRA30 expression cultures.

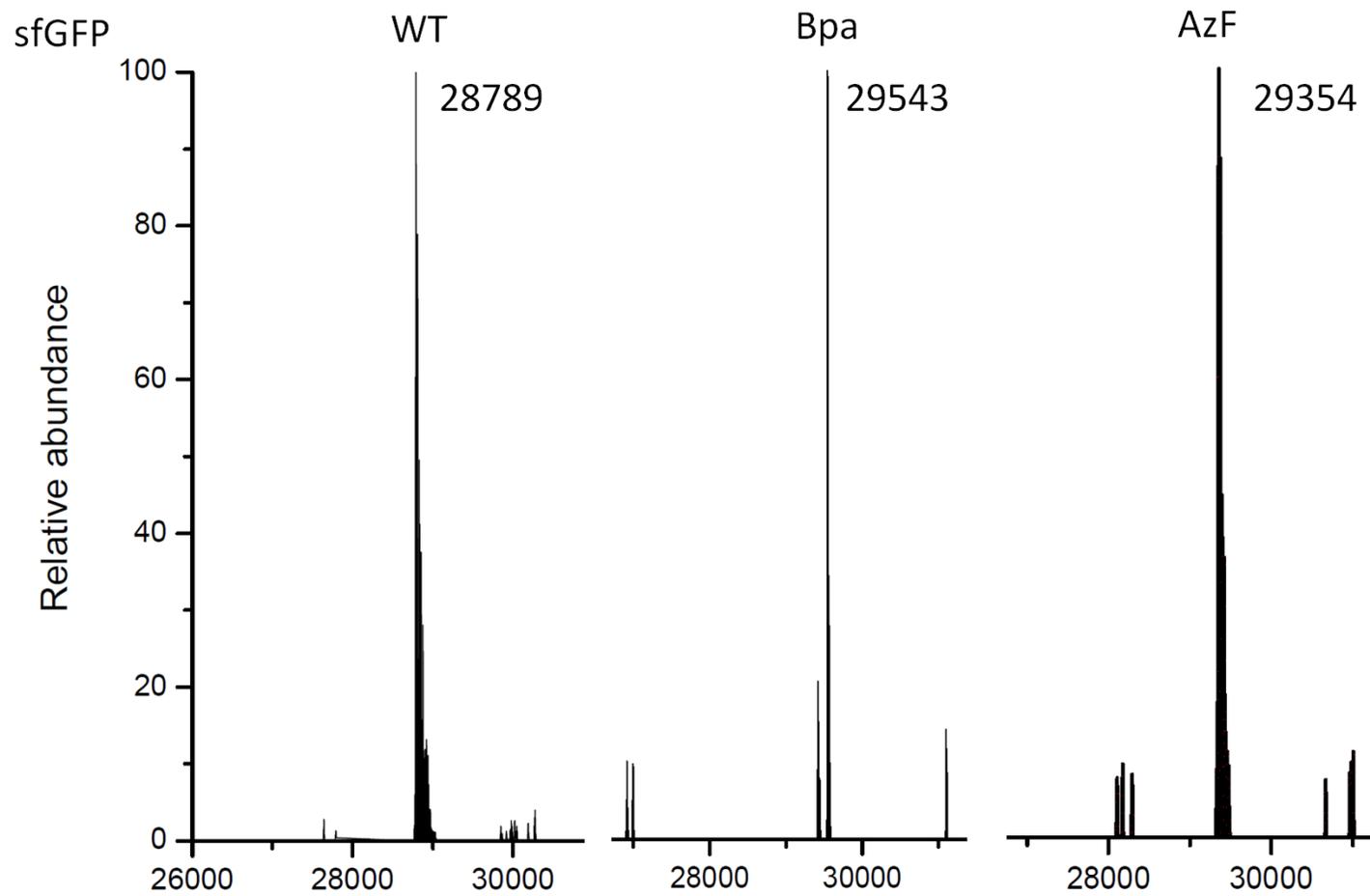


Figure 10. Deconvoluted ESI-mass spectra of purified wild-type sfGFP, sfGFP-(AzF)₃, and sfGFP-(Bpa)₃.

Table 3. Superfolder GFP expression and ESI-mass spectrometry data.

sfGFP	Yield (mg/L) ^[a]	Theoretical m/z (u) ^[b]	Experimental m/z (u)	(Δ WT) _{exp} (u) ^[c]	(Δ WT) _{calc} (u) ^[d]
WT	60.8	28791.3	28789	-	-
Bpa	4.1	29545.2	29543	756	754
AzF	13.4	29355.9	29354	563	565

[a] Yield was determined by the Bradford assay on purified protein solutions.

[b] Theoretical molar mass was calculated based on the assumption that *E. coli* methionyl-aminopeptidase results in proteolytic cleavage of the *N*-terminal methionine to afford the (sfGFP–Met) derivatives.

[c] The difference in experimental molar masses between the substituted sfGFP-(Xaa)₃ and wild-type sfGFP.

[d] Calculated molar mass for three equivalents of the indicated non-canonical amino acid as a peptidic unit (calcd molar mass: Bpa, C₁₆H₁₃NO₂, 251.28 u; AzF, 188.19 u.)

Elastin-mimetic peptides with photo-crosslinkers

We are interested in the ability to employ this system to support biosynthetic incorporation of the photo-crosslinkers into elastin-mimetic protein. The elastin-mimetic polypeptides utilized here were described earlier in Chapter 2. Briefly, the protein polymer encodes amino acid sequence of 12 or 22 repeats of the pentapeptide monomer, [(Val-Pro-Gly-Val-Gly)₂(Val-Pro-Gly-Amber-Gly)(Val-Pro-Gly-Val-Gly)₂]. We also attempted to increase the overall yield of elastin derivatives by expression in a richer media, 2YT or Terrific Broth (TB), instead of the LB medium (Data not shown). Western blot analyses indicated the successful expression of full-length elastin-(UAG)₁₂ and elastin-(UAG)₂₂ in the *E. coli* MRA30 strain only in the presence of respective analogues (Fig. 11). In addition, the elastin-(UAG)_n derivatives could be purified via metal affinity chromatography with significant yield. As observed earlier of the sfGFP derivatives, the yields of the AzF-substituted elastin-mimetic polypeptides were higher (16.8 mg/mL and 10 mg/mL for the 12 and 22 repeat sequences, respectively) than that of corresponding Bpa-substituted proteins (8.3 mg/mL and 6.4 mg/mL). However, the yields reported here for the incorporation of these photo-crosslinkers were quite respectable, especially considering the high density of the amber codons encoded in the target elastin-mimetic polypeptide sequences.

The identities of full-length elastin-(UAG)_n polymer were confirmed via MALDI-TOF mass spectrometry. The molecular ions of elastin-(Bpa)₁₂ and elastin-(AzF)₁₂ present in the spectra were consistent with the calculated molecular weight with a high level of substitution into the corresponding elastin sequence (Fig. 12 and Table 4). However, the longer elastin derivatives, elastin-(UAG)₂₂, were not detected in either case of Bpa or

AzF analogues, presumably due to low desorption/ionization efficiency of the higher molecular weight elastin polymers substituted with these hydrophobic phenylalanine derivatives. In addition, in order to ascertain the site-specific location of the analogues within the polypeptides, protease digestion and tandem mass spectrometry were employed. Thermolysin was chosen due to its broad specificity for the digestion of the elastin polymers that could be resulted in the sizeable fragments suitable for mass spectrometry analysis. The fragments digested from elastin-(Bpa)₁₂ present in the ESI-spectrum were relative dirty with the presence of contaminated polymer signals. The source of the polymers might come from surfactants, plastics, tubing, etc, during the purification of elastin-(Bpa)₁₂ under denaturing condition. However, the molecular ion of the [VGVPG(Bpa)GVPG]⁺ at m/z 989 that showed mass differences attributed to the presence of Bpa analogues was observed (Fig. 13). These parent ions were selected for tandem mass analysis in order to recover the sequence information. Furthermore, MADLI mass spectrometry, which has a higher tolerance for impure samples, was also employed to verify the sequence of fragments that resulted from proteolysis (Fig. 14).

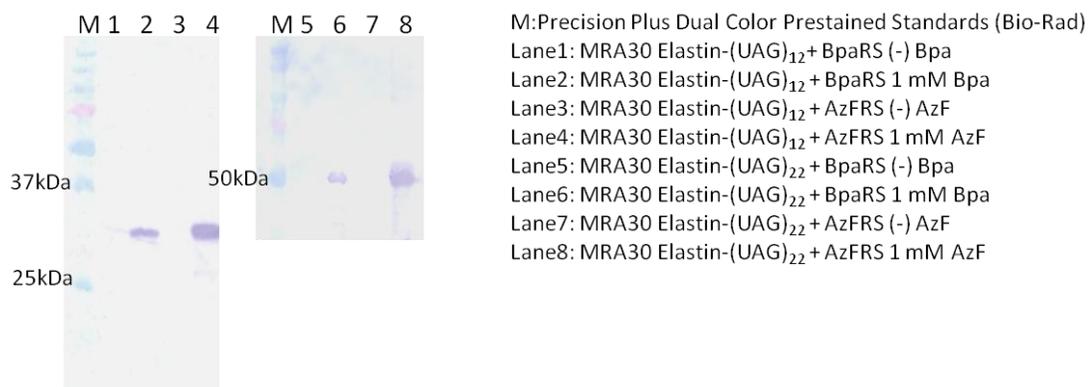
Digestion of elastin-(AzF)₁₂ resulted in shorter peptides compared to Bpa-substituted analogues. It is possibly due to the structural similarity between AzF and the cognate tyrosine; the latter of which is an active cleavage site for thermolysin. The ESI-MS/MS spectrum of the molecular ions at m/z 616 presented the b-/y-ions corresponding to the sequence of VGVPG(AzF) (Fig. 15). These mass spectra suggest that the photocrosslinkers, Bpa and AzF, could be repetitively and site-specifically localized within the polypeptide sequence of the elastin polymers in response to the amber codons. The full-length proteins could be purified from unoptimized expression cultures of the RF1-

attenuated *E. coli* MRA30 strain in sufficient yield for subsequent experimental investigations of crosslinking efficiency.

Table 4. Elastin protein expression and MALDI-TOF mass spectrometry data.

Elastin ^[a]	Yield [mg/L] ^[a]	Theoretical m/z ^{[b],[c]}	Experimental m/z
(Bpa) ₁₂	8.3	28984	28964
(Bpa) ₂₂	6.4	n.d.	n.d.
(AzF) ₁₂	16.8	28228	28240
(AzF) ₂₂	10.0	n.d.	n.d.
<p>[a] Elastin derivatives with the Bpa and AzF residues were purified via immobilized metal affinity chromatography and the yield was measured by dry weight.</p> <p>[b] Molar masses for the elastin derivatives were calculated based on the complete substitution with the non-canonical amino acid at the amber-encoded sites. The calculated molar mass assumes proteolytic cleavage of the <i>N</i>-terminal methionine residue as a consequence of the endogenous activity of the <i>E. coli</i> methionyl-aminopeptidase.</p> <p>[c] Theoretical molar mass was based on the amino acid sequence of the elastin polypeptides: [(VPGVG)₂VPGXG(VPGVG)₂]_nVPGVGGSDDDDKGH₁₀ as [M+H]⁺ species for the Bpa and AzF derivatives.</p>			

(A)



(B)

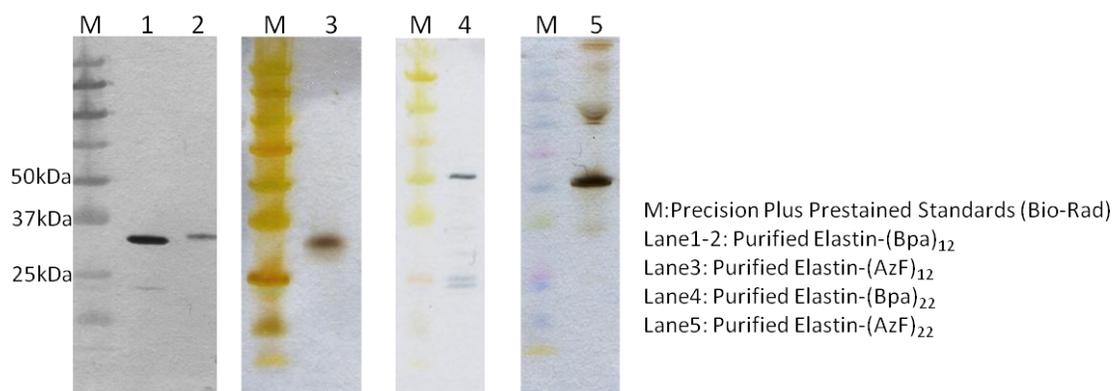
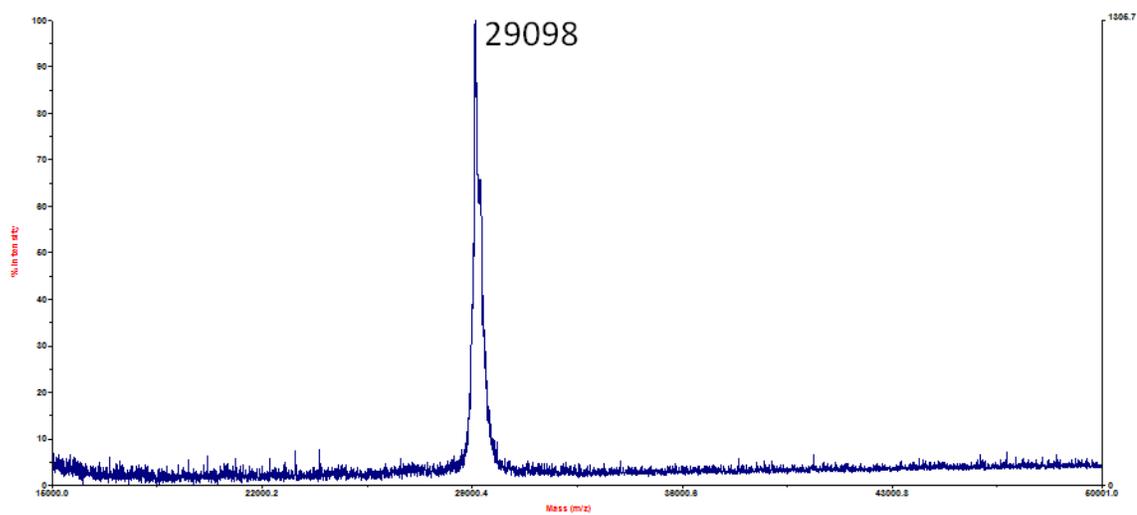


Figure 11. (A) Western blot analyses of the expression of elastin-(Bpa)_n and elastin-(AzF)_n derivatives in *E. coli* MRA30 strain. (B) SDS-PAGE analyses of the purified elastin-(Bpa)₁₂, elastin-(Bpa)₂₂, elastin-(AzF)₁₂, and elastin-(AzF)₂₂.

(A)



(B)

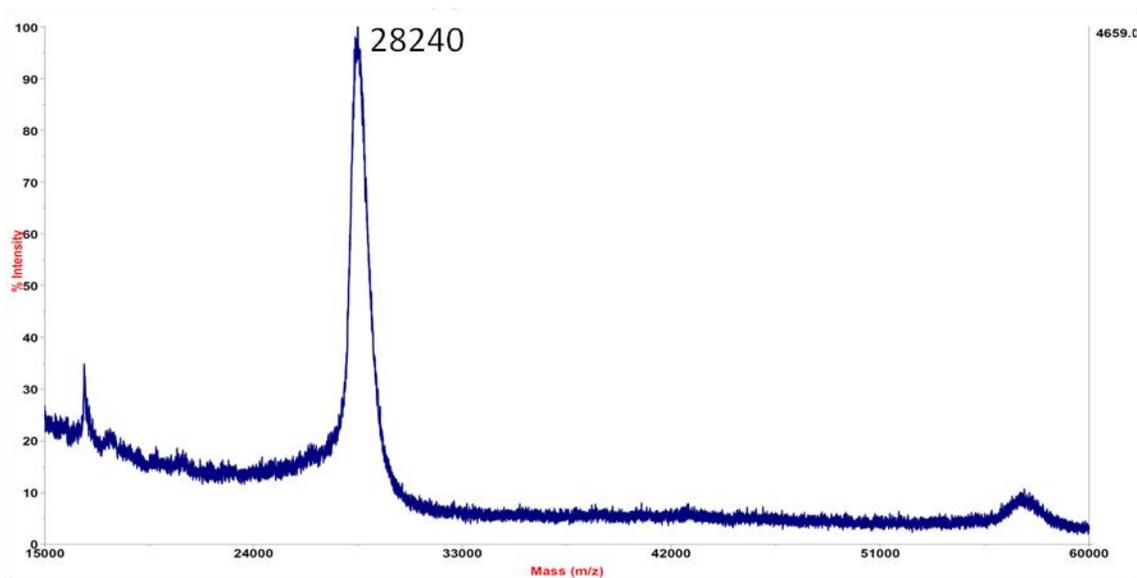
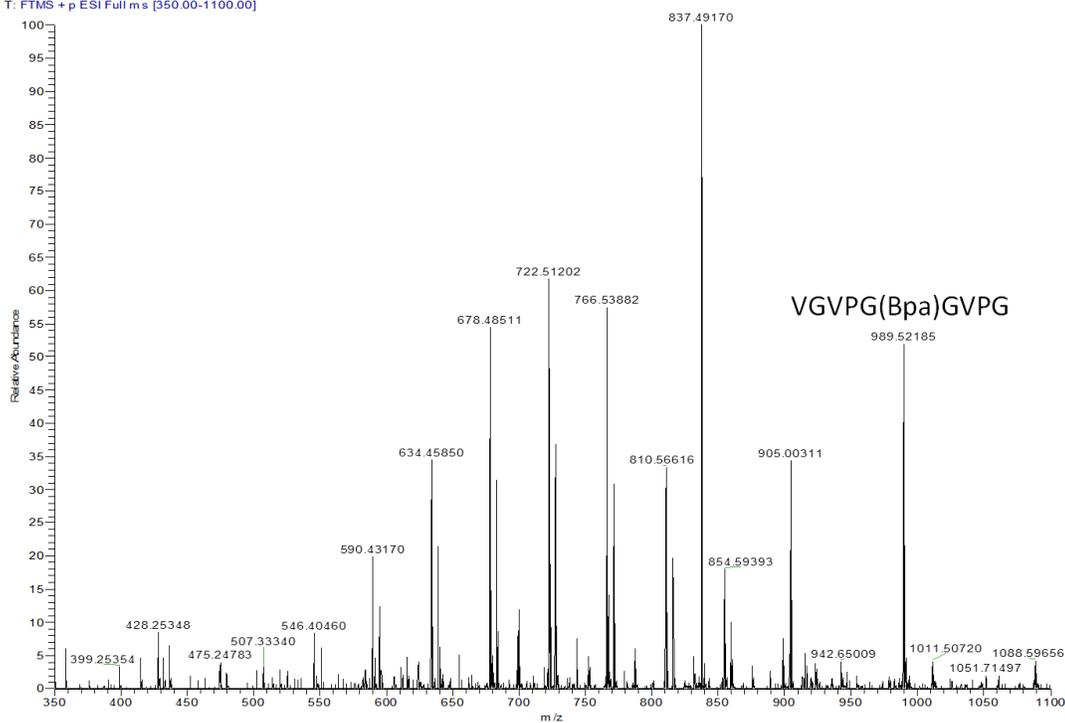


Figure 12. MALDI-TOF mass spectra of the full-length elastin derivative, elastin-(Bpa)₁₂ (A) and elastin-(AzF)₁₂.

(A)

FT17548_110705153219#1 RT: 0.00 AV: 1 NL: 4.76E4
T: FTMS + p ESI Full ms [350.00-1100.00]



(B)

FT17548_110705153501#1-4 RT: 0.00-0.08 AV: 4 NL: 1.54E5
T: FTMS + p ESI Full ms 2 989.52@cid30.00 [270.00-1100.00]

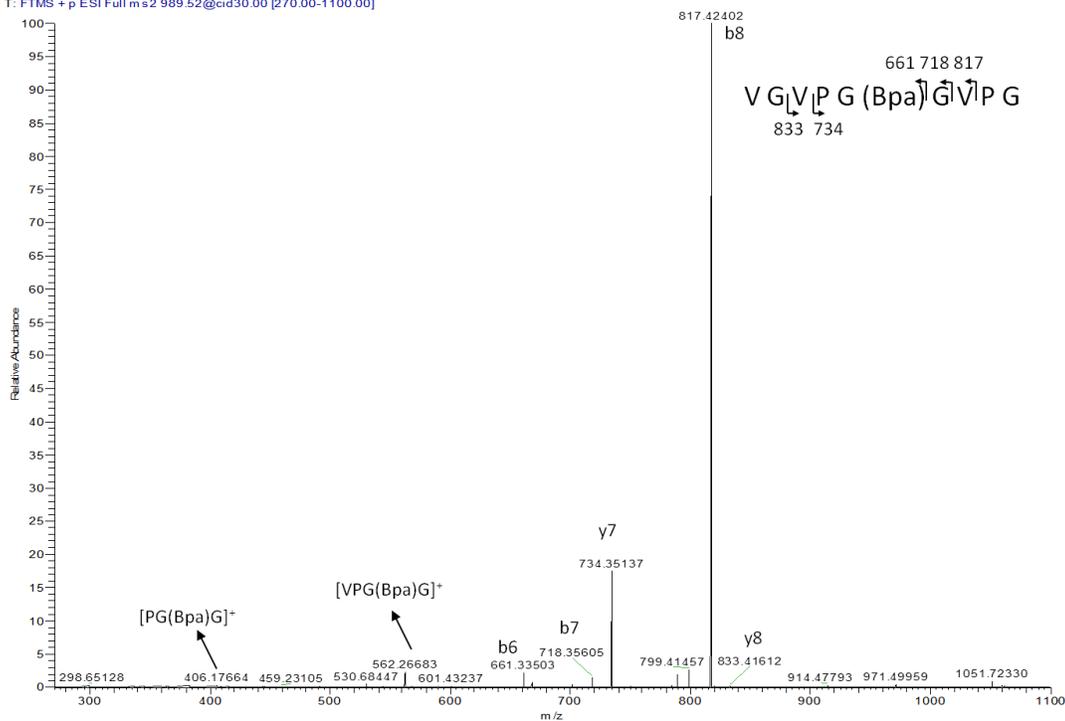


Figure 13. ESI-MS (A) and MS/MS (B) spectra of elastin-(Bpa)₁₂ derivatives digested by thermolysin.

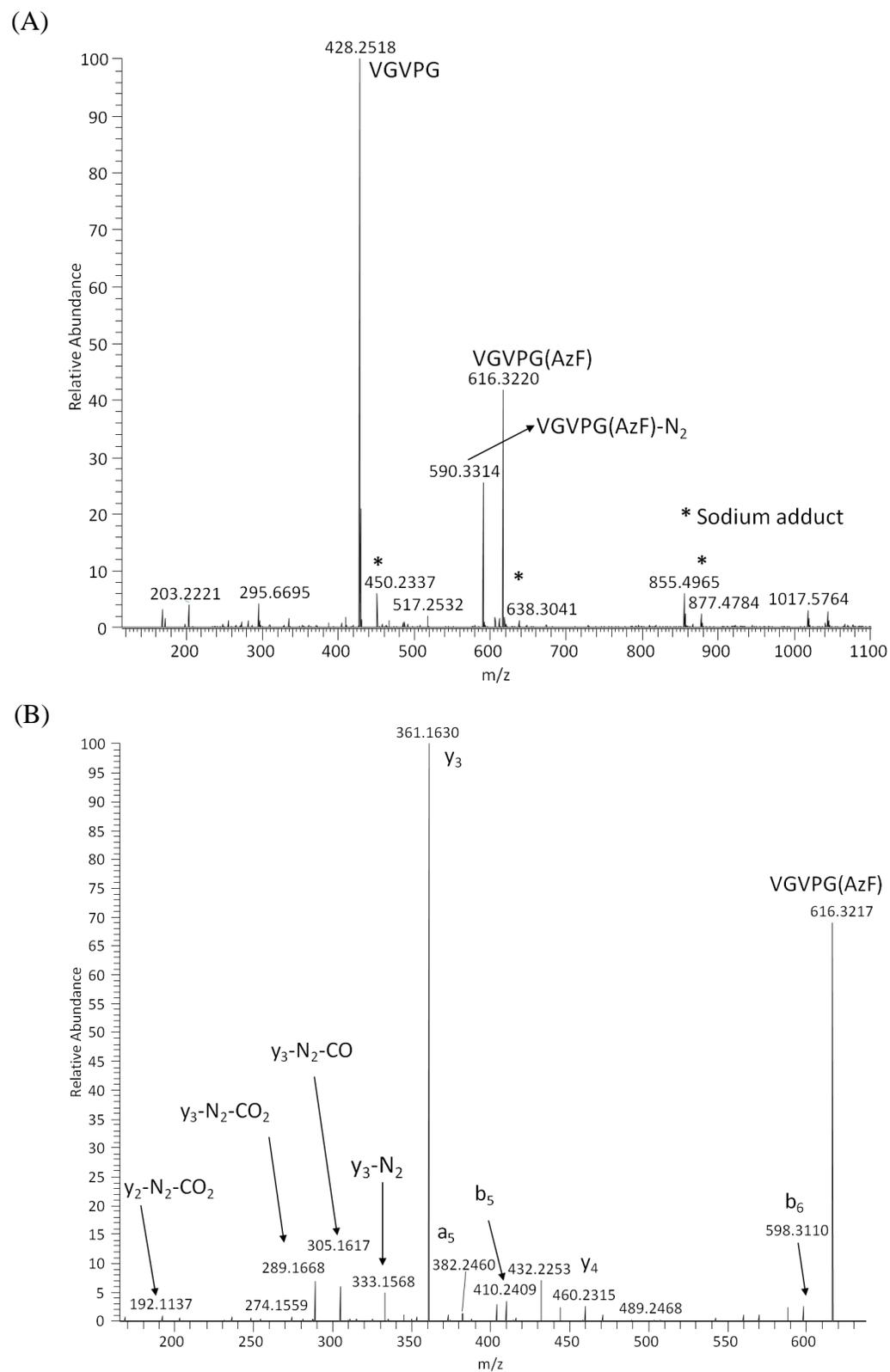


Figure 15. ESI-MS (A) and MS/MS (B) spectra of elastin-(AzF)₁₂ derivatives digested by thermolysin.

Photo-crosslinkable activity

Since the AzF derivatives afforded higher yields and demonstrated more tractable behavior in solution, these derivatives were employed as substrates to investigate the feasibility of UV-induced cross-linking of proteins that had been multiply labeled. We compared the protein properties before and after UV exposure (365 nm, 30 min) by electrophoresis and Western blotting. Upon the UV radiation, the model protein sfGFP-AzF was converted to a covalently linked multimers (Fig. 16). In contrast, control experiments for wild-type sfGFP showed no detectable difference in response to UV irradiation. These results demonstrate that the incorporated AzFs could be adapted as photo-probes initiated by UV exposure and beneficial for mapping protein-protein interactions. The goal for the labeling application is to introduce unique amino acids at particular sites that would not affect the function, but would be informative with respect to the protein conformational changes in order to study the assembly or the structural dynamics of supramolecular assemblies. The multi-site specific incorporation system demonstrated here should facilitate the study of protein functions or interactions by designing or performing multiple labeling using bio-orthogonal photo-crosslinkers.

Elastic-mimetic polymer with AzF analogues display inverse phase transition near the room temperature that can be monitored via turbidimetric analysis. Since elastin polymers might form turbid coacervate above the transition temperature^{46,47}, model photo-crosslinking experiments were carried out on the elastin-(AzF)₁₂ at 4 °C and 37 °C respectively. Gel electrophoresis showed that the cross-linked coacervate at 37 °C was not able to move into the separation gel possibly due to the formation of large aggregate after crosslinking (Fig. 17). However, the experiments performed at 4 °C showed a range

of crosslinked proteins which might result from the crosslinking of the proximal soluble elastin-(AzF)₁₂ unimers at low temperature. Therefore, our results demonstrated that the newly synthesized elastin-mimetic polypeptides may serve as a useful module for the generation of novel biomaterials, for example, amphiphilic block copolymers, capable of photo-crosslinking.

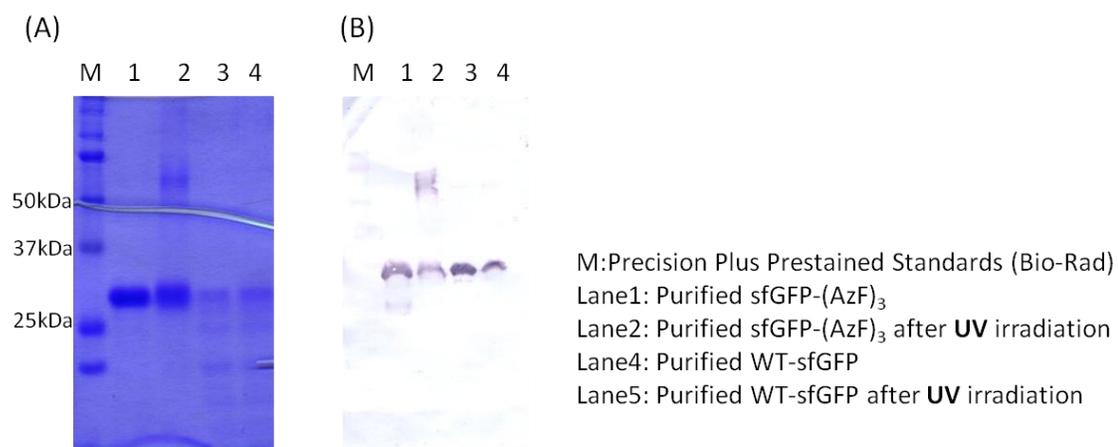


Figure 16. (A) SDS-PAGE and (B) Western blot analysis of sfGFPs in response to UV irradiation.

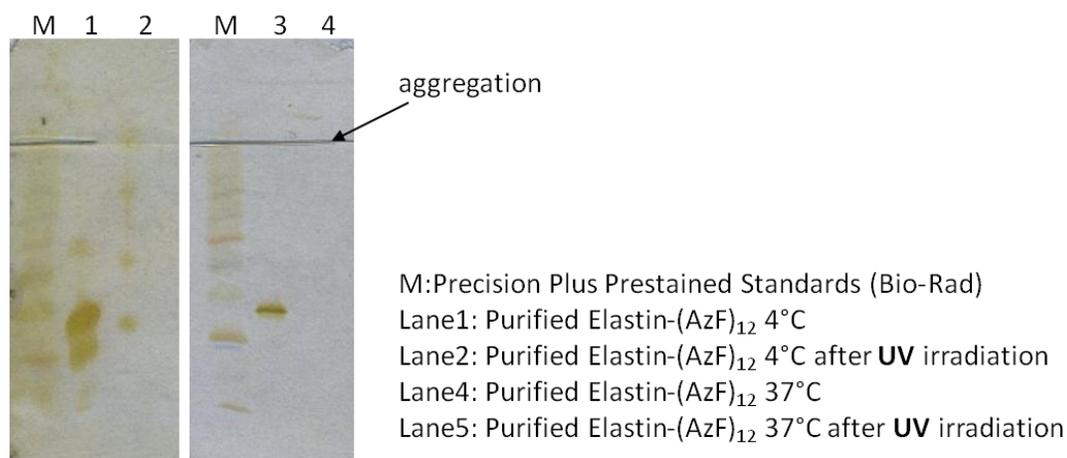


Figure 17. SDS-PAGE analysis of elastin-(AzF)₁₂ in response to UV irradiation.

Conclusion

The application of photo-affinity probes in biological systems suffers from a few technical limitations. In comparison with the conventional semi-synthetic approach, which allows the incorporation of a wide variety of UAAs into peptides, the orthogonal aaRS/suppressor tRNA pair strategy has significantly enhanced the technical convenience.⁴⁸ In this chapter, we optimize an expression system that allows the *in vivo* synthesis of functionally active proteins containing photo-initiated probes at multiple-specific sites. We adapted the amber suppression strategy to genetically incorporate the photocrosslinkable UAAs, Bpa or AzF, into proteins expressed in the modified *E. Coli* strain with attenuated release factor-1 activity. We have utilized evolved MjtRNA variants with significantly enhanced activity for the incorporation of these analogues into target proteins. The activity of BpaRS was improved by active-site mutations and the evolved BpaRS pair was permissive to Bpa analogues with five-fold higher activity compared to the parent pair. The MjtRNA Nap1 showed enhanced activity for both Bpa and AzF synthetase. This system reported here is compatible with most *E. coli* expression vectors and thus is useful for the design and high-yield expression of mutant proteins with selectively substituted photo-crosslinkers that are genetically designated by amber termination codons.

Reference

- [1] D. W. Urry, C.-H. Luan, C. M. Harris, T. M. Parker, in *Protein-Based Materials* (Eds.: D. Kaplan, K. McGrath), Birkhauser, Boston, **1997**, pp. 133-177.
- [2] V. P. Conticello, H. E. Carpenter Desai, in *Polymer Science: A Comprehensive Reference Vol. 9* (Eds.: K. Matyjaszewski, M. Möller), Elsevier, **2012**, pp. 71-103.
- [3] S. R. MacEwan, A. Chilkoti, *Biopolymers* **2010**, *94*, 60-77.
- [4] H. M. Kagan, S. N. Gacheru, P. C. Trackman, S. D. Calaman, F. T. Greenaway, **1989**, 317-326.
- [5] B. Vrhovski, A. S. Weiss, *Eur. J. Biochem.* **1998**, *258*, 1-18.
- [6] K. Reiser, R. J. McCormick, R. B. Rucker, *FASEB J.* **1992**, *6*, 2439-2449.
- [7] E. R. Wright, R. A. McMillan, A. Cooper, R. P. Apkarian, V. P. Conticello, *Adv. Funct. Mater.* **2002**, *12*, 149-154.
- [8] X. Wu, R. Sallach, C. A. Haller, J. A. Caves, K. Nagapudi, V. P. Conticello, M. E. Levenston, E. L. Chaikof, *Biomacromolecules* **2005**, *6*, 3037-3044.
- [9] R. E. Sallach, W. Cui, J. Wen, A. Martinez, V. P. Conticello, E. L. Chaikof, *Biomaterials* **2009**, *30*, 409-422.
- [10] S. Bozzini, L. Giuliano, L. Altomare, P. Petrini, A. Bandiera, M. T. Conconi, S. Farè, M. C. Tanzi, *J. Mater. Sci.: Mater. Med.* **2011**, *22*, 2641-2650.
- [11] G. T. Hermanson, *Bioconjugate Techniques, 2nd Edition*, Elsevier, **2008**.
- [12] J. Gubbens, E. Ruijter, L. E. V. de Fays, J. M. A. Damen, B. de Kruijff, M. Slijper, D. T. S. Rijkers, R. M. J. Liskamp, A. I. P. M. de Kroon, *Chem. Biol.* **2009**, *16*, 3-14.

- [13] S. Ye, T. Huber, R. Vogel, T. P. Sakmar, *Nat. Chem. Biol.* **2009**, *5*, 397-399.
- [14] A. Grunbeck, T. Huber, P. Sachdev, T. P. Sakmar, *Biochemistry* **2011**, *50*, 3411-3413.
- [15] N. Hino, Y. Okazaki, T. Kobayashi, A. Hayashi, K. Sakamoto, S. Yokoyama, *Nat. Methods* **2005**, *2*, 201-206.
- [16] G. Dorman, G. D. Prestwich, *Biochemistry* **1994**, *33*, 5661-5673.
- [17] J. C. Kauer, S. Erickson-Viitanen, H. R. Wolfe, Jr., W. F. DeGrado, *J. Biol. Chem.* **1986**, *261*, 10695-10700.
- [18] G. W. J. Fleet, R. R. Porter, J. R. Knowles, *Nature* **1969**, *224*, 511-512.
- [19] Y. Chen, Y. Ebricht, R. Ebricht, *Science* **1994**, *265*, 90-92.
- [20] L. K. Dyllal, J. E. Kemp, *Aust. J. Chem.* **1967**, *20*, 1395.
- [21] T. Huber, S. Naganathan, H. Tian, S. Ye, T. P. Sakmar, **2013**, *520*, 281-305.
- [22] S. S. van Berkel, M. B. van Eldijk, J. C. M. van Hest, *Angew. Chem. Int. Ed.* **2011**, *50*, 8806-8827.
- [23] S. Milles, S. Tyagi, N. Banterle, C. Koehler, V. VanDelinder, T. Plass, A. P. Neal, E. A. Lemke, *J. Am. Chem. Soc.* **2012**, *134*, 5187-5195.
- [24] R. L. M. Teeuwen, S. S. van Berkel, T. H. H. van Dulmen, S. Schoffelen, S. A. Meeuwissen, H. Zuilhof, F. A. de Wolf, J. C. M. van Hest, *Chem. Commun.* **2009**, 4022.
- [25] K. Zhang, M. R. Diehl, D. A. Tirrell, *J. Am. Chem. Soc.* **2005**, *127*, 10136-10137.
- [26] I. S. Carrico, S. A. Maskarinec, S. C. Heilshorn, M. L. Mock, J. C. Liu, P. J. Nowatzki, C. Franck, G. Ravichandran, D. A. Tirrell, *J. Am. Chem. Soc.* **2007**, *129*, 4874-4875.

- [27] P. J. Nowatzki, C. Franck, S. A. Maskarinec, G. Ravichandran, D. A. Tirrell, *Macromolecules* **2008**, *41*, 1839-1845.
- [28] J. Raphel, A. Parisi-Amon, S. C. Heilshorn, *J. Mater. Chem.* **2012**, *22*, 19429-19437.
- [29] J. W. Chin, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020-11024.
- [30] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026-9027.
- [31] S. Ye, M. Riou, S. Carvalho, P. Paoletti, *ChemBioChem* **2013**, *14*, 230-235.
- [32] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413-444.
- [33] T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, *395*, 361-374.
- [34] I. S. Farrell, R. Toroney, J. L. Hazen, R. A. Mehl, J. W. Chin, *Nat. Methods* **2005**, *2*, 377-384.
- [35] A. L. Stokes, S. J. Miyake-Stoner, J. C. Peeler, D. P. Nguyen, R. P. Hammer, R. A. Mehl, *Mol. BioSyst.* **2009**, *5*, 1032.
- [36] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883-890.
- [37] T. Dale, R. P. Fahlman, M. Olejniczak, O. C. Uhlenbeck, *Nucleic Acids Res.* **2008**, *37*, 1202-1210.
- [38] T. Dale, L. E. Sanderson, O. C. Uhlenbeck, *Biochemistry* **2004**, *43*, 6159-6166.
- [39] Y. Doi, T. Ohtsuki, Y. Shimizu, T. Ueda, M. Sisido, *J. Am. Chem. Soc.* **2007**, *129*, 14458-14462.
- [40] J. Guo, C. E. Melancon, 3rd, H. S. Lee, D. Groff, P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9148-9151.

- [41] S. Zhang, M. Ryden-Aulin, L. A. Kirsebom, L. A. Isaksson, *J. Mol. Biol.* **1994**, *242*, 614-618.
- [42] K. Gabriel, W. H. McClain, *J. Mol. Biol.* **2001**, *310*, 543-548.
- [43] H. E. Carpenter Desai, Emory University **2008**.
- [44] M. A. Patterson, Emory University **2011**.
- [45] L. Bossi, *J. Mol. Biol.* **1983**, *164*, 73-87.
- [46] D. W. Urry, C. H. Luan, T. M. Parker, D. C. Gowda, K. U. Prasad, M. C. Reid, A. Safavy, *J. Am. Chem. Soc.* **1991**, *113*, 4346-4348.
- [47] D. W. Urry, D. C. Gowda, T. M. Parker, C. H. Luan, M. C. Reid, C. M. Harris, A. Pattanaik, R. D. Harris, *Biopolymers* **1992**, *32*, 1243-1250.
- [48] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413-444.

Chapter 4

System application: Synthesis of photo-crosslinkable elastin diblock copolymers

Introduction

Biomaterials derived from protein-based blocks are increasingly investigated for their potential applications in biology and medicine.¹⁻⁴ Many of the applications envisage the fabrication of advanced nano-carriers to deliver the cargo when they reach the target cells or tissue. Stimuli-responsive nano-carriers have been developed to release drugs into tumor tissues upon changes in physical and chemical environments, such as redox potential, pH, temperature, and ultrasound.⁵⁻⁸ Polymeric micelles have been widely researched as one of the most promising nano-carriers for water-insoluble anticancer drugs.⁹⁻¹¹ Normally, a core-shell micellar structure is achieved from the self-assembly of amphiphilic copolymers in the aqueous solutions. Those well-developed nanoparticles exhibit a prolonged circulation time by avoiding the rapid clearance from the reticulo-endothelial system as a consequence of the small size of the particles and the perfect hydrophilicity and hydratability of the corona consisting of poly(ethylene glycol) (PEG) or dextran.^{12,13} However, conventional polymeric micelles employed as drug carriers suffer from the drawback of disaggregation when diluted into body fluids, giving rise to premature release of drugs. Recently, genetically-directed synthesis has been employed to design protein block copolymers with uniform sequences from simple amino acid building blocks with an absolute degree of control and complexity superior to those achieved by synthetic copolymers.¹⁴⁻¹⁶ The advantages, such as good biodegradability, ease of preparation and controllable functionality, make protein amphiphilic copolymers a promising class of polymers for biomedical applications.^{10,14,17} In most cases, reversible protein self-assembly could be driven by the well-defined conformational changes of peptide units induced in response to an external stimulus. Protein amphiphilic diblock

copolymers that incorporate targeting ligands, cell-adhesion sequences, or fluorescent groups are of significant technological interest, especially in conjunction with the ability to integrate stimuli-responsive ability in response to the change in the surroundings.¹⁵

In particular, recombinant elastin block copolymers provide attractive characteristics as substrates for the creation of novel responsive material architectures in that the phase transition can be controlled through stimuli such as temperature, pH, redox state and other environmental factors.^{6,7,14,18} Elastin-mimetic polypeptides based on the pentameric repeat sequence [Val-Pro-Gly-Xaa-Gly] have been demonstrated to undergo thermo- and pH-responsive self-assembly in aqueous solution.¹⁹⁻²¹ Spontaneous phase separation occurs due to a conformational rearrangement of local secondary structure above a unique transition temperature (T_t) that depends on the chemical identity of the fourth amino acid (Xaa) in the pentapeptide repeat.²⁰ For example, the elastin-mimetic blocks that contain hydrophobic amino acids such as tyrosine in the fourth position display a conformational transition from random coils to repetitive type II β -turns at temperature below 37 °C, whereas blocks containing a charged amino acid persist as a random coil throughout the physiologic temperature range.²² Thus, a wide range of amphiphilic block copolymers could be designed through the combination of different hydrophobic and hydrophilic peptide structures derived from the native protein sequence.² Selective segregation of the hydrophobic block induced by temperature occurs in aqueous solutions promoting micelle-like aggregates formation in which the corona is derived from the hydrophilic block B and the core from the hydrophobic block A (Fig. 1).²³ In addition, the exploration of tissue development and regeneration has led to the discovery of a multitude of cell-instructive peptide domains, which can be included in an

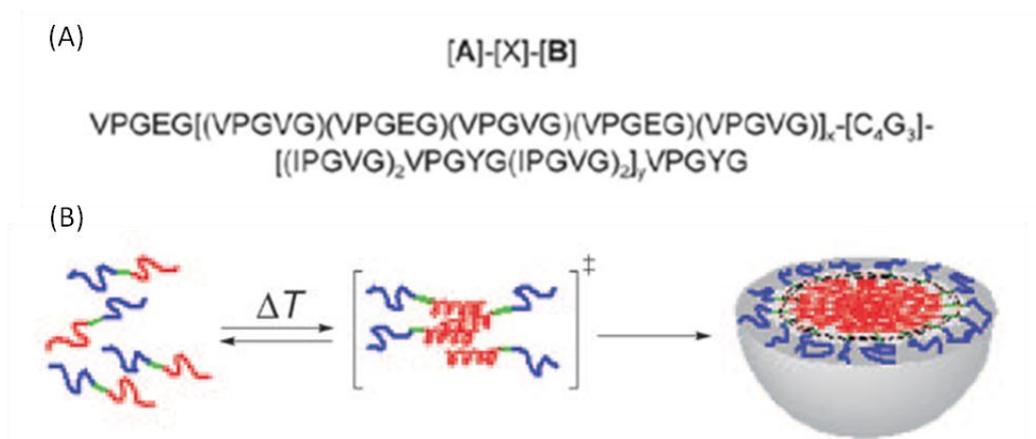


Figure 2. (A) Sequence and (B) schematic of disulfide cross-linked elastin diblock copolymers.¹⁸

application.^{29,30} Many research groups have developed various crosslinking approaches to enhance micelles stability, however, only a few reports have been presented on the development of cross-linked biodegradable micelles for applications in controlled release, eg., anticancer drug delivery. Chaikof and co-workers recently demonstrated engineered thermally responsive and crosslink-stabilized protein nanoparticles produced by the tailored design of recombinant elastin amphiphilic diblock copolymers.¹⁸ They showed that the positioning of cysteine residues between the elastin blocks could avoid uncontrolled micelle-micelle association and stabilize nanoparticles through the intermolecular covalent disulfide crosslinking at the core-shell interface (Fig. 2).¹⁸ Furthermore, these stabilized nanoparticles with disulfide crosslinks would limited the release of the hydrophobic drug from the protein core; an attribute that is useful for prolonged delivery of fluorescent molecules or drugs. However, the disulfide crosslinks have several potential disadvantages that include susceptibility to over-oxidation, disulfide exchange, and dis-assembly under reducing conditions.

Most protein-based sequence-repetitive biomaterials are synthesized by utilizing the translational machinery of *E. coli* to encode the desired amino acid sequence. Since the exact protein sequence is specified and reliably produced through recombinant methods, these materials offer exquisite control over the resulting properties.³¹ In addition to the possibility to include bioactive and cell-instructive peptide domains, specific amino acids can be positioned at defined locations for chemical modifications. For example, cysteine residues are often employed for thiol reactions, while lysine residues are targeted for primary amine chemistries.³² These sites are frequently used to attach other molecules such as drugs or fluorescent dyes to yield bioactive protein-conjugates or to react with crosslinking molecules for the formation of bulk biomaterials. However, crosslinking reagents might have potential cytotoxicity and result in inhomogeneity with residual reaction byproducts. Photo-crosslinking is a fast processing method commonly used to overcome these concerns in the production of a variety of biomedical synthetic polymeric materials.³³ This technique confers the ability to easily pattern polymer structures, which have wide-ranging applications in the development of drug delivery vehicles, implant coatings, and tissue engineering scaffolds. Furthermore, the use of focused light sources and photo-masks could facilitate exquisite spatial and temporal control of crosslinking reactions. While photoreactions are a common strategy for processing synthetic polymers, proteins are not inherently photoactive. Therefore, the addition of photoactive functionality into protein-based biomaterials is particularly useful.

Protein-based biomaterials with photo-crosslinkers integrated have combined a rapid, controllable, and biocompatible crosslinking method with a modular genetic design strategy to create a new family of bioactive materials.³⁴⁻³⁶ Heilshorn and coworkers

demonstrated the successful derivation of a bioactive, recombinant engineered elastin-like polypeptide (ELP) with a photoactive diazirine moiety to enable the versatile photo-processing of stable, cell-adhesive biomaterials.³⁷ The bioactivity of this ELP material could promote cell adhesion, along with its versatility to produce coatings, thin films, and bulk scaffolds followed by the crosslinking through UV exposure. As an alternative strategy, photo-reactivity could be conferred on the protein through the incorporation of non-canonical, photo-reactive amino acids directly into the protein sequence. Previous work in Tirrell's group has demonstrated the incorporation of the non-canonical amino acid, *para*-azido-phenylalanine (AzF), into engineered ELP materials that has designed as artificial extracellular matrix proteins included a cell-adhesive domain.³⁴ The ELP polymers were expressed in the auxotrophic bacterial hosts with minimal medium incapable of synthesizing phenylalanine, which is then replaced by the non-canonical analogue, AzF. The extent of phenylalanine replacement varied from 13 to 66%, depending on the concentration of the analogues in the expression medium. Their work demonstrated that photo-crosslinkable ELPs can be fabricated into 2D and 3D networks that still retain their relevant bioactive functionalities for cell adhesion. The wide range of processing possibilities, high stability, biocompatibility, and modular bioactivity has made photo-crosslinkable ELP materials a promising choice for further development of diverse biomedical applications. However, there are certain key limitations affecting the efficacy of these residue-specific systems while the appeal of using naturally derived materials for *in vivo* applications is convenient. Protein yields using genetically engineered auxotrophic hosts in combination with the defined minimal medium are typically low, limiting the scalability and widespread application of this approach. In

addition, these materials might suffer from significant batch-to-batch variations in their mechanical and biochemical properties due to the incomplete global substitution by amino acid analogues, leading to irreproducible performance. In previous chapters of this study, we have demonstrated the system for the multiple site-specific incorporation of unnatural amino acids co-translationally into elastin-mimetic polypeptides in *E. coli*. In addition, we have also demonstrated the efficient and site-selective incorporation of the photo-reactive amino acids into the elastin polymers, retaining their functionality for photo-crosslinking.

In this chapter, we utilize the system that we previously developed for the generation of novel photo-crosslinkable amphiphilic elastin block copolymers that could self-assemble into core-shell-corona nanoparticles. Previous studies in our lab have demonstrated the ability of an amphiphilic diblock copolymer derived from the elastin-mimetic peptide sequences to undergo reversible, temperature-dependent segregation in aqueous solution affording potentially biocompatible nanoparticles under physiologically relevant conditions.¹⁴ Therefore, the installation of the photo-initiated functionality within the amphiphilic block copolymers could provide the potential to capture features unique to both physical and photo-crosslinking schemes and subsequently permit the control over a wide range of chemical, mechanical, and biological properties afforded by the protein-based multi-domain materials. We have constructed an amphiphilic (**AB**-type) diblock copolymer that incorporates previously synthesized hydrophilic elastin concatemer (**S1**), [(VPGAG)₂VPGE(VPGAG)₂]₁₉³⁸ and the elastin-UAG domain. We hypothesize that the elastin-UAG domain will exhibit a lower inverse temperature transition due to the presence of non-polar phenylalanine derivatives at the amber

suppression sites. Basically, the amber suppression strategy enables the incorporation of diverse functionalities into the elastin-UAG domain within the **AB**-type diblock and provides new avenues for the design of novel amphiphilic elastin block copolymers. The advantages of the photo-crosslinker, AzF, have been described in the previous chapter and it is therefore selected in this study for the design of photo-crosslinkable elastin diblock copolymers. These AB-type amphiphilic block copolymers should self-assemble into micellar nanoparticles in the aqueous solution above the transition temperature of the elastin-(AzF) block. In addition, facile photo-crosslinking of the micelles should result from UV irradiation to form core-crosslinked nanoparticles. We report herein the novel thermally responsive amphiphilic elastin diblock copolymers that form covalently photo-crosslinked, stabilized biocompatible nanoparticles. Moreover, this novel photo-crosslinked micellar system is expected to have the potential as drug carriers for controlled release. The **AB**-type elastin diblock copolymers generated in this work demonstrate that the incorporation of unnatural amino acids with unique functional groups can confer novel chemical properties that enhance the functional performance of protein-based materials.

Experimental Methods

Materials

Chemical reagents were purchased from Fisher Scientific, Inc. (Pittsburgh, PA) or Sigma Chemical Co. (St. Louis, MO) unless otherwise noted. The amino acid derivative *para*-azido-*L*-phenylalanine (AzF) was purchased from Bachem Bioscience, Inc. (King of Prussia, PA). Isopropyl- β -D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs, Inc. (Beverly, MA). The *E. coli* strain TOP10F' was obtained from Invitrogen Corp. (Carlsbad, CA). The *E. coli* strain MRA30 was kindly provided by permission of Dr. Monica Rydén-Aulin with the assistance of Jaroslav Belotserkovsky.³⁹ Plasmid pZErO[®]-1 and pZErO[®]-2 were obtained from Invitrogen Corp. (Carlsbad, CA).

General Methods

Basic molecular biology techniques, including ligation, immunoblotting, and gel electrophoresis, were adapted from a standard molecular cloning manual 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 (0.22 μ m 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. Plasmid DNA preparation and DNA purification were performed using the QIAprep Spin Miniprep Kit (QIAGEN Inc., Valencia, CA) and DNA Clean & Concentrator[™] (Zymogen, CA). Ligations were carried out in a 10 μ L volume with 0.5

μ L T4DNA Ligase (200 units) and 1X T4DNA Ligase buffer at 16 °C overnight. Chemically competent cells of TOP10F' were transformed with ligation mixtures and recovered at 37 °C for an hour. The recovery mixture (200 μ L) was spread onto LB agar supplemented with appropriate antibiotics. The plates were incubated at 37 °C for 12 to 14 hours. Proteins were purified by metal-affinity chromatography using the cobalt-charged TALON[®] Metal affinity Resin from Clontech Laboratories (Mountainview, CA). The plasmids and *E. coli* strains utilized and constructed in the synthesis of elastin diblock copolymers are summarized in Table 1.

Construction of the elastin-diblock plasmids

The gene encoding the ELP-S1 hydrophilic block (pSCP3.11) was available from a previous study.³⁸ A plasmid was generated that encodes a diblock (**AB**-block) copolymer where the previously synthesized hydrophobic, elastin-UAG block is encoded at the N-terminus with the hydrophilic S1 block located at the C-terminus. The directional cloning strategy required that both DNA concatemer cassettes be encoded in the same cloning plasmid (Fig. 3).³¹ Since the antibiotic resistance gene for the S1 and elastin-UAG multimers cloning plasmids pZErO[®]-2 (Fig. 4) and pZErO[®]-1 differ, the elastin-UAG multimer was excised from the pZErO[®]-1 vector via the digestion of pIL42.12 with *Bam*H I and *Hind* III, and ligated into compatible sites within the pZErO[®]-2 plasmid. Recombinant plasmids were screened by double digestion with *Bam*H I and *Hind* III to identify the elastin-UAG multimer gene in the plasmid. The purified recombinant plasmid clones were sequenced by using M13 reverse and M13 (-20) forward primers, and designated as pIL103.12 (Fig. 5).

The pIL103.12 and pSCP3.11 cloning plasmids, which contain the elastin-UAG

and S1 concatemers respectively, were utilized for the directional assembly of a diblock plasmid. A ligation was performed between the *BsmB I* / *Nco I*-digested pIL103.12 (~80 ng) and *Bbs I* / *Nco I*-digested pSCP3.11 (~100 ng) in a 1:1 molar ratio. The ligation mixture was used to transform TOP10F' cells, which were plated onto low salt LB agar supplemented with kanamycin (50 µg/mL). The recombinant plasmids were screened by the digestion with *BamH I* and *Hind III* and DNA fragments were analyzed by agarose gel electrophoresis for the presence of the diblock insert. The sequence of the purified plasmid DNA, designated as pIL105, was confirmed by automated DNA sequencing analysis with M13 reverse and M13 (-20) forward primers for the presence of a gene fusion of the elastin-UAG multimers and S1 block flanked by *Bbs I* and *BsmB I* restriction endonuclease recognition sites (Fig. 6). Next, an acceptor plasmid suitable for protein expression was required to conduct the high-level expression of the diblock in a bacterial host. The pZER0[®]-2 plasmid, pIL105, containing the corresponding diblock (~3000 bp) concatemers was sequentially digested with *Bbs I* and *BsmB I* to isolate the multimer genes. The pIL5 plasmid was digested with *Bsa I* to generate the complimentary sticky ends necessary for the seamless cloning of the elastin-AB diblock genes into the adaptor sequence of the modified pQE-80L expression plasmid. The digested pIL5 DNA was dephosphorylated by antartic phosphatase to reduce the false positive ligations. Ligation mixture was transformed into TOP10F' and spread on LB agar supplemented with ampicillin (100 µg/mL). The recombinant plasmid DNA was isolated from the cultured colonies and screened via the double digestion with *EcoR I* and *Hind III* for the presence of the diblock (**AB**-type) concatemer inserts. The sequence of the expression plasmids was confirmed by automated DNA sequencing analysis using

pQE-F/ R primers. The resulting plasmid was designated as pIL106 that includes a modified pQE80L and the ELP-diblock copolymer under control of the phage T5 promoter inducible by IPTG (Fig. 6).

Protein expression and purification

Chemically competent cells of *E. coli* strain MRA30 were prepared. Cells were transformed with the plasmid pIL106 (elastin-AB diblock), and selected on LB agar supplemented with ampicillin (100 µg/mL) via incubation at 30 °C for 12-14 h. Fresh electro-competent cells were prepared for further transformation. The plasmid pIL101 containing the optimized MjAzFRS pair was transformed and plated on LB agar media supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) to generate expression strains for these studies.

Single colonies of the expression strains were inoculated into sterile LB broth supplemented with the appropriate antibiotics (100 mg/mL ampicillin and 34 mg/mL chloramphenicol) as required for plasmid maintenance. One liter of TB supplemented with the appropriate antibiotics was inoculated with 20 mL of overnight culture and incubated at 30 °C until OD₆₀₀ reached between 0.6 and 0.8 absorbance units. Amino acid analogues (AzF) and IPTG were added as a final concentration of 1 mM to induce the expression of the elastin derivatives. After 20 h, the cells were harvested by the centrifugation at 4000 g and 4 °C for 20 min. The cell pellet was re-suspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and stored at -80°C.

The frozen cells were lysed by three freeze/ thaw cycles between -80°C and room temperature. Lysozyme (1 mg/mL), EDTA-free protease inhibitor cocktail, benzonase (25 units/mL), and MgCl₂ (1 mM) were added to the lysate and the mixture was

incubated shaking at 4 °C for 12 h. The cell lysate was centrifuged at 14,000 g for 30 min at 4 °C. The supernatant from the soluble elastin-mimetic proteins with AzF analogues was loaded directly onto a column containing cobalt-charged TALON resin (3 mL) and washed with lysis buffer (20 mL) containing 20 mM imidazole. The target protein was eluted with elution buffer (20 mL, 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0) and dialyzed (MWCO = 10 kDa) against distilled de-ionized water (5 x 4 L). The dialysate of elastin-mimetic protein was lyophilized to produce a white, spongy solid and the yield was measured by dry weight.

Transmission electron microscopy

TEM was used to identify the potential nanometer-sized structures that resulted from the self-assembly. Dilute aqueous solutions of ELP-AB diblock copolymer (1 mg/mL) were prepared at pH 6.0 (10 mM MES buffer), pH 8.5 (10 mM TAPS buffer), and pH 10.0 (CAPS buffer). The solutions were incubated at 45 °C for approximately 1 h to induce the hydrophobic assembly of the elastin-UAG domain and generate micellar nanoparticles. Photo-crosslinking was performed via UV radiation using 356nm hand-held UV lamp for 30 min. Aliquots (4 µL) of the polymer solutions were then deposited onto the surface of a 200 mesh carbon film-coated copper grids at room temperature and allowed to sit for 60 s. Excess solvent was wicked away and 4 µL of the negative stains Nano-W[®] (methylamine tungstate) from Nanoprobes (Yaphank, NY) was added onto the grids. 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 JEOL 1210 TEM at a 70 kV accelerating voltage.

Temperature-dependent turbidity

The thermal transition of the elastin diblock copolymers was estimated from temperature turbidity profiles. Measurements were performed using a JASCO V-530 UV/Visible spectrophotometer equipped with a programmable Peltier cell and a water-controlled JASCO ETC-505T temperature controller (Jasco, Inc.; Easton, MD). Dilute solutions of the elastin-mimetic peptide samples (1.0 mg/mL) were prepared at 4 °C and analyzed using a quartz cuvette with a 10 mm path-length (Hellma Analytics; Müllheim, Germany). The absorbance of the polymer solutions was monitored over a temperature range from 4 °C to 60 °C at a 600 nm wavelength with a ramp rate of 1 °C/min. Rescan of the samples was performed *in situ* following by 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.

Dynamic light scattering

Dilute aqueous solutions of ELP-AB diblock copolymers (1 mg/mL) were prepared in the aqueous solution at pH 8.5 (10mM TAPS buffer). The solutions were incubated at 45 °C for approximately 1 h to induce the hydrophobic assembly of the hydrophobic domain chains and generate nanoparticles. DLS experiments were performed using Zetasizer Nano ZS90 (Malvern, Worcestershire, UK) instrument equipped with a 633nm red laser. The accessible scattering angle is fixed at 90 degree. The samples were stabilized at the desired temperature for 10 min prior to the measurement. The data were acquired and the average correlation function was analyzed by Zetasizer® software.

Atomic force microscopy

Dilute aqueous solutions of ELP-AB diblock copolymers (1 mg/mL) were prepared at pH 8.5 (10mM TAPS buffer). The solutions were incubated at 45 °C for approximately 1 h to induce the hydrophobic assembly of the elastin-UAG domain to generate micellar nanoparticles. Photo-crosslinking was performed via UV radiation using 356nm hand-held UV lamp for 30 min. Morphology information of photo-crosslinked nanoparticles on pre-cleaned mica coverslips was recorded on the JEOL JSPM-4200 microscope. Images were processed and analyzed using WSxM. Silicon AFM tips (MikronMasch) with a force constant (5.4-16 N/m) were used in the tapping mode at a scan rate of 1 Hz. SiO₂ grating TGZ-20 was used for z-axis calibration (Ted Pella).

Thermolysin digestion and mass spectrometry

Dithiothreitol (DTT) was added into the ELP-AzF diblock copolymers as a final concentration of 10 mM and the mixture was incubated at 100 °C for 30 min to denature the protein. Thermolysin was added to the denatured protein solution to a concentration of 1: 50 (w/w) ratio with respect to protein and the reaction mixture was incubated at 37 °C for 12 h. The products from the proteolysis reaction were passed through a PepCleanTM C18 spin column (Thermo Scientific, Inc.) to remove the salts and thermolysin. MALDI spectra of thermolysin-digested peptides were carried out using Applied Biosystem 4700 Proteomics mass spectrometer. A solution of the α -cyano-4-hydroxycinnamic acid (4-CHCA) in 50 % ACN and 0.2% FA at a concentration of 10 mg/ml was used as matrix. The peptide sample was mixed with the matrix solution in a ratio of 1:20 and deposited onto the MALDI plate. Mass spectra were acquired at the range of m/z 500 to 2500 in the positive reflectron mode.

Table 1. Plasmids and strains utilized in Chapter 4.

Plasmids	Relevant characteristics	Reference
pZero [®] -2	<i>E. coli</i> cloning vector, Kan ^R	Invitrogen
pIL101	MjAzFRS/Nap1 tRNA pair in pHEC2, Cm ^R	Chapter 3
pIL5	pIL2 with removal of internal <i>Bsa</i> I, Amp ^R	Chapter 2
pIL42.12	Elastin-UAG _{14mer} concatemer in pZero [®] -1, Zeo ^R	This study
pIL103.12	Elastin-UAG _{14mer} concatemer in pZero [®] -2, Kan ^R	This study
pSCP3.11	S1 concatemer in pZero [®] -2, Kan ^R	Reference ³⁸
pIL105	Elastin-UAG _{14mer} concatemer and S1 concatemer (Elastin-diblock) in pZero [®] -2, Kan ^R	This study
pIL106	Elastin-UAG _{14mer} concatemer and S1 concatemer (Elastin-diblock) in pIL5, Amp ^R	This study
Strain	Genotype	Reference
Top10F'	F [<i>lacIq</i> , Tn10(TetR)] <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) Φ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ (<i>ara leu</i>) 7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i>	Invitrogen
MRA30	<i>prfA1 recA rph</i>	Reference ³⁹

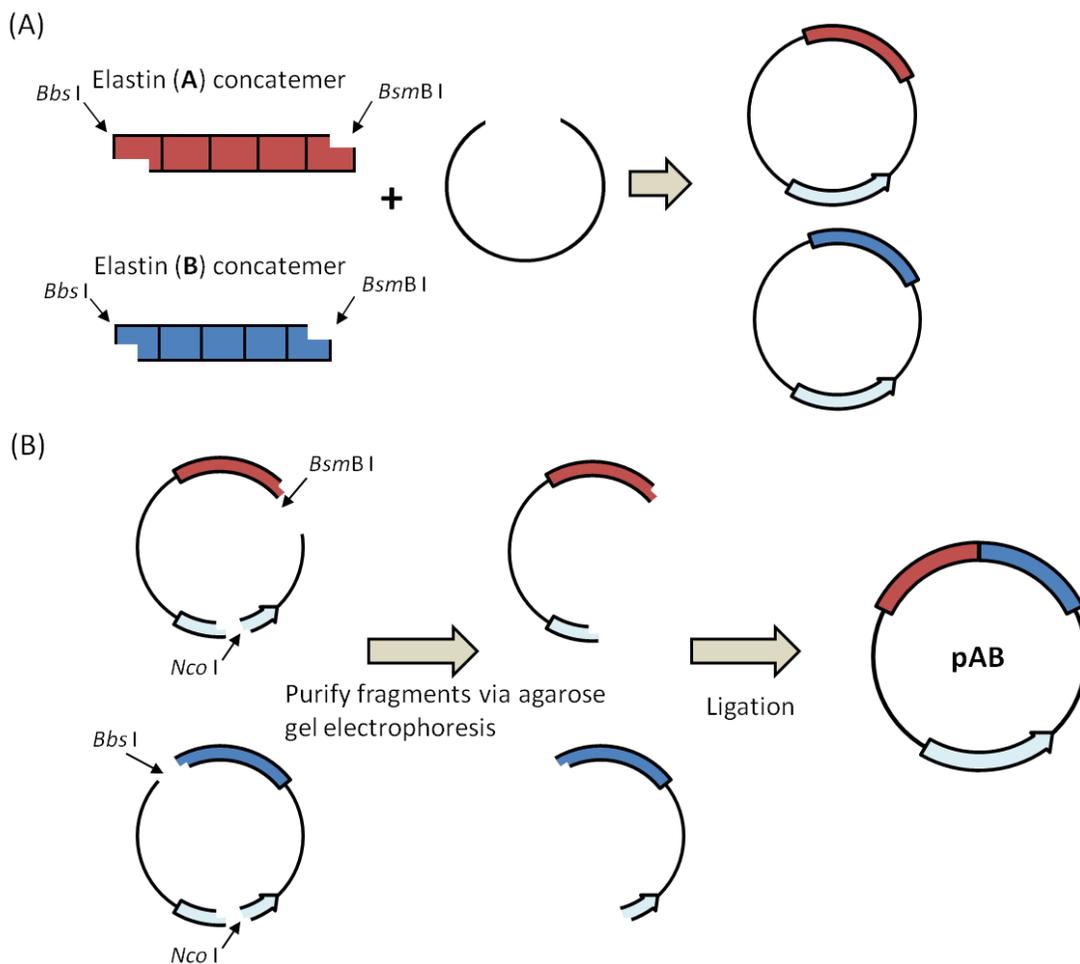
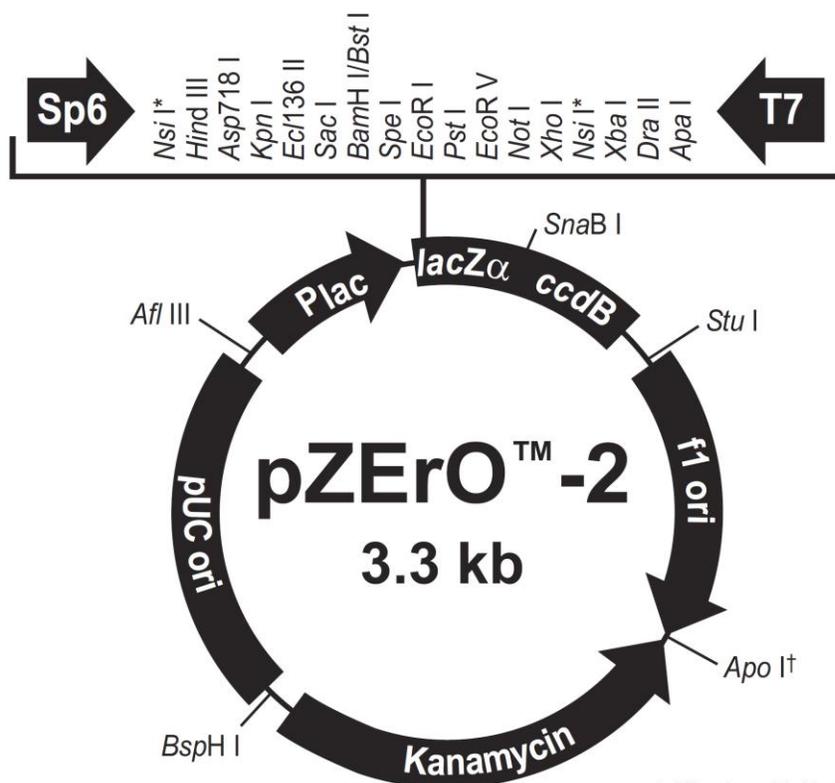


Figure 3. Directional cloning strategy for assembly of synthetic genes encoding diblock polypeptides. (A) Plasmid clones encoding the elastin-A and elastin-B concatemers are cloned into the same type cloning plasmid, typically pZErO[®]-2 or pZErO[®]-1 plasmids. (B) The DNA concatemer plasmids are then assembled to create an **AB**-diblock or **BA**-diblock gene fusion.



Comments for pZErO™-2
3297 nucleotides

Lac Promoter/Operator Region: bases 95-216
 M13 Reverse Priming Site: bases 205-221
LacZα ORF: bases 217-558
 Sp6 Promoter/Priming Site: bases 239-256
 Multiple Cloning Site: bases 269-381
 T7 Promoter/Priming Site: bases 388-407
 M13 (-20) Forward Priming Site: bases 415-430
 M13 (-40) Forward Priming Site: bases 434-450
 Fusion Joint: bases 559-567
ccdB Lethal Gene ORF: bases 568-870
 f1 origin: bases 895-1307
 Kanamycin Resistance ORF: bases 2116-1322 (C)
 pUC origin: bases 2502-3175

* The two *Nsi* I sites in the MCS are the only sites in the vector.

† There are two tandem *Apo* I sites at this location. *Apo* I also recognizes the *EcoR* I site.

Figure 4. Plasmid map of the pZErO[®]-2 vector, from Invitrogen, Inc., indicating restriction endonuclease cleavage sites within the multiple cloning site. The **Elastin-S1** DNA monomer was cloned into the plasmid using the *BamH* I/ *Hind* III sites.

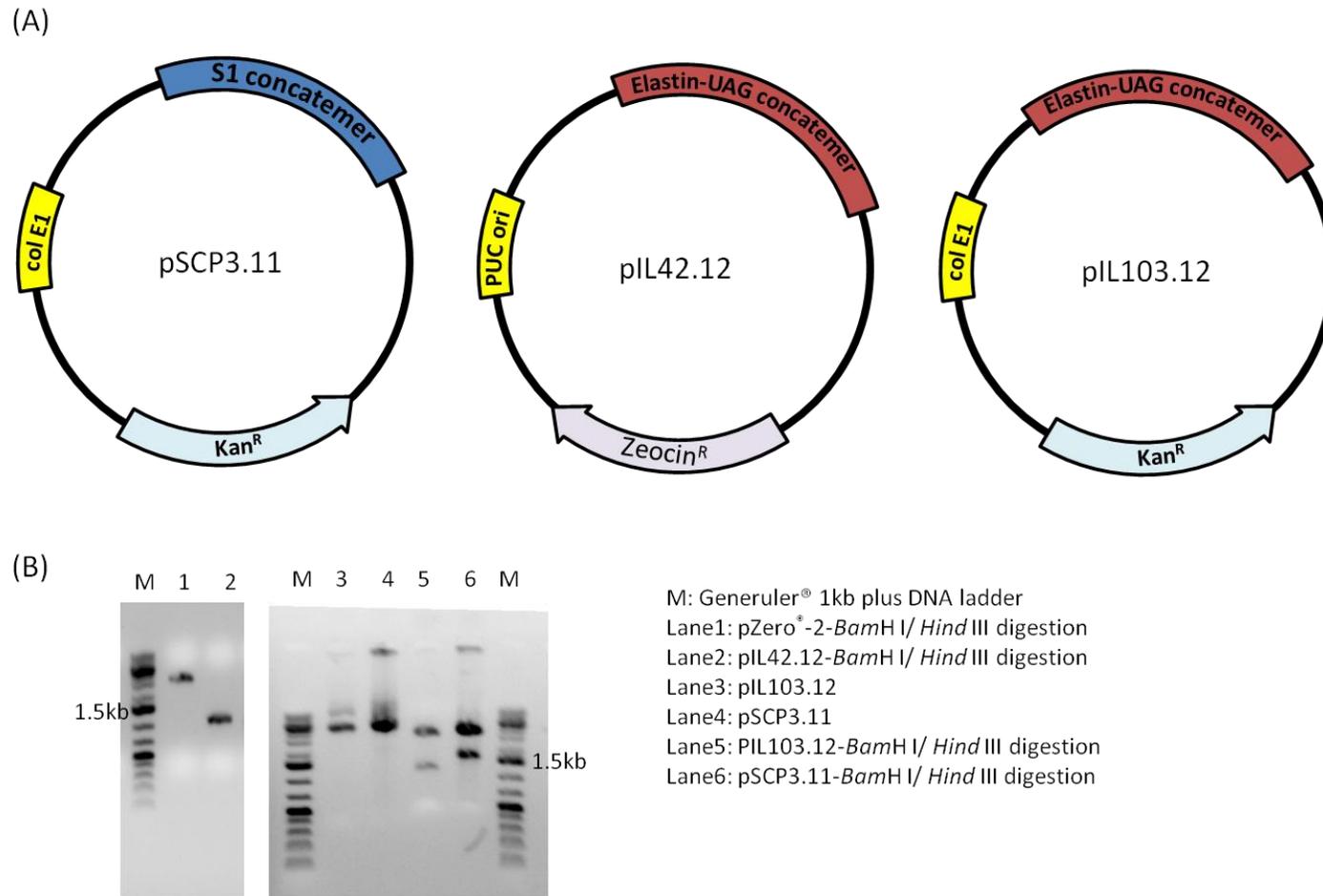


Figure 5. (A) Plasmids pSCP3.11 and pIL103.12 encoding elastin-S1 and elastin-UAG_{14mer} concatemers in pZerO®-2. Plasmid pIL42.12 encoding elastin-UAG concatemers in pZerO®-1. (B) Agarose gel electrophoresis the digestion of pSCP3.11, pIL42.12 and pIL103.12.

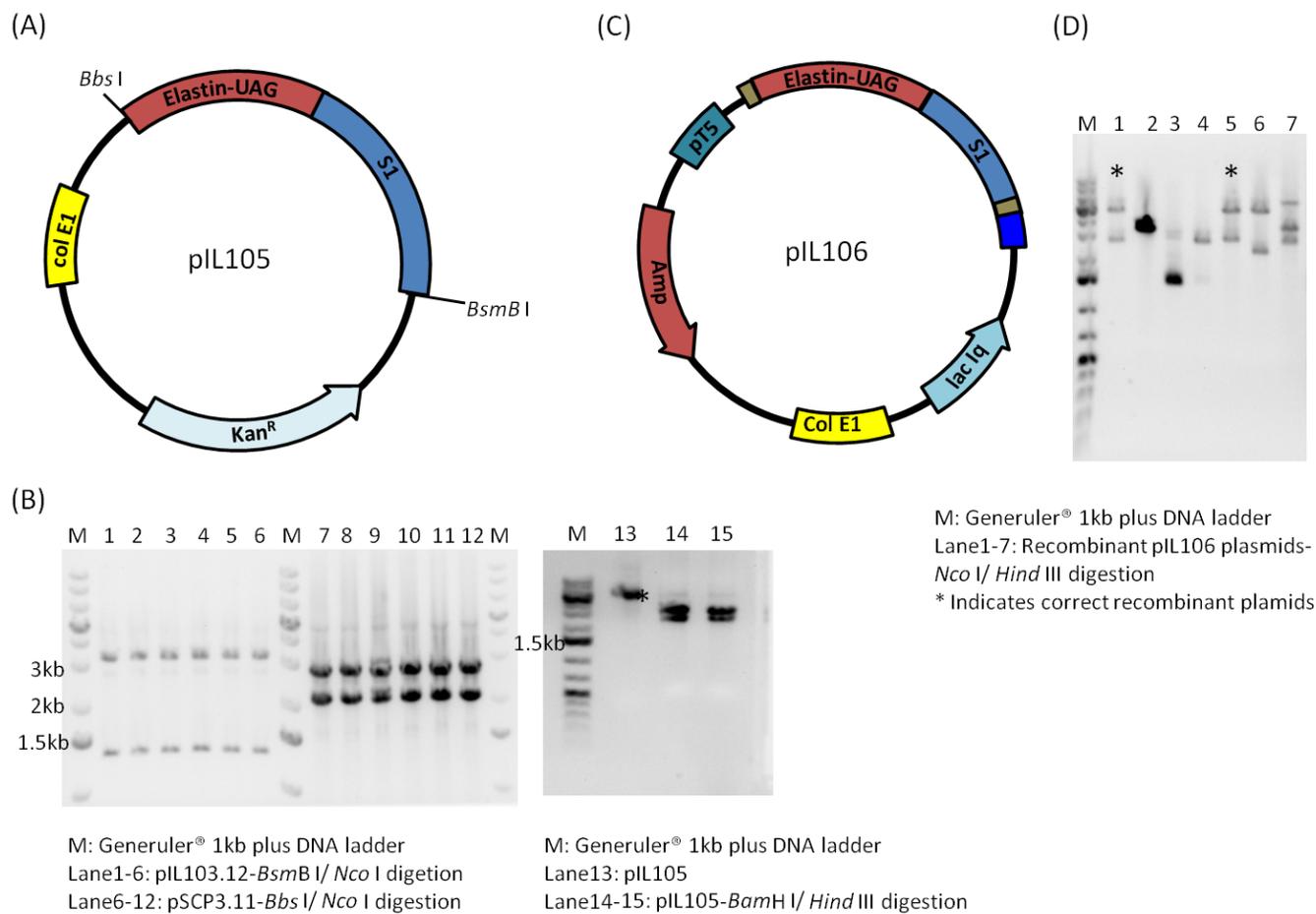


Figure 6. (A) Plasmid pIL105 encoding diblock gene fusion in pZErO[®]-2. (B) Agarose gel electrophoresis the digestion of pSCP3.11, pIL103.12 and pIL105. (C) plasmid pIL106 encoding the diblock polymer in the expression plasmid pIL5. (D) Agarose gel electrophoresis of pIL106 recombinant plasmids.

Results and Discussion

The self-assembly of amphiphilic diblock copolymers results in the formation of a spherical micelle consisting of a core made up of an insoluble block, which is shielded from the solvent by a hydrated corona composed of a more soluble block.¹⁴ The dimensions of these nanoparticles should facilitate long-term circulation. We report herein thermally responsive elastin-mimetic polypeptides in the linear AB diblock architecture that can self-assemble into spherical micelles with the ability to maintain structural integrity via photo-crosslinking upon UV irradiation. The photocrosslinkable amino acid, AzF, is incorporated at periodic sites within the ELP via co-translational insertion in response to amber sites located in the gene sequence of the *N*-terminal hydrophobic block. Transmission electron microscopy (TEM) and atomic force microscopy (AFM) are utilized to monitor the formation and morphology of these nanoparticles. The temperature responses of the protein nanoparticles are investigated by the dynamic light scattering (DLS). The photo-crosslinked ELPs remain as spherical nanoparticles unresponsive to further thermal variations. This unique property could be applicable for the design of drug carriers to achieve prolonged drug release and enhanced tumor accumulation. Moreover, this available genetic platform further allows the adaptation of a variety of amber suppression strategy using orthogonal pairs to generate ELP block copolymers with the exquisite design of selective non-natural amino acids.

Design and production of Elastin-UAG/S1 (ELP-AB) diblock copolymer

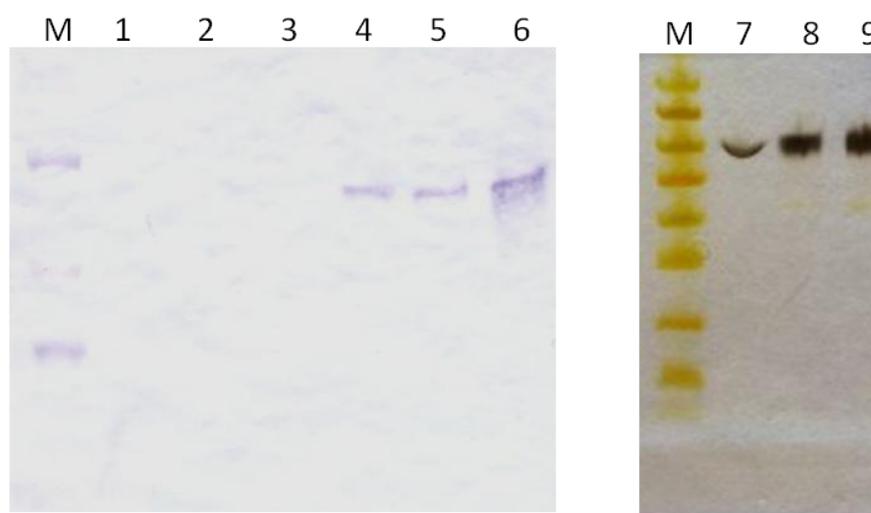
The AB-type amphiphilic ELP block copolymers were designed using previously reported directional assembly strategy (Fig. 3).³¹ The hydrophobic and hydrophilic blocks are encoded within separate DNA cassettes that had been concatemerized independently

and joined together to create the genetic fusion of the ELP-AB diblock. This strategy involves the cleavage of the DNA concatemer cloning plasmids with a type II restriction endonuclease and an enzyme, which cleaves within the antibiotic resistance gene of the cloning plasmid. The DNA fragments are purified and ligated to produce a genetic fusion of the proximal and distal concatemers. Digestion within the antibiotic-resistance gene reduces the incidence of false positives during the subsequent selection of recombinant clones since the ligation of the fragments reconstitutes the antibiotic marker. The ELP diblock gene was cloned into the pQE-80L-modified expression plasmid subsequently with the tunable elastin-UAG_{14mer} block (14mer) at the N-terminus and the hydrophilic elastin-S1 block (19mer) at the C-terminus (Fig. 6).

The ELP diblock copolymer was co-expressed with the MjAzF orthogonal pair in the RF1-attenuated *E. coli* MRA30 strain. Western blot results showed the time-course expression of the ELP diblock copolymers in the TB medium supplemented with AzF analogues (Fig. 7). The ELP-AzF diblock migrated to a position that corresponded to a slightly larger molecular weight than those of the calculated value, a trend that was reported previously.^{15,40} In the SDS-PAGE, ELPs may have an extended conformation due to the random coil, which could result in a slower migration rate than the globular protein standard having the same size. The ELP-AzF copolymers were isolated from large scale cultures using metal affinity purification with yields approximately 17 mg/L and the purity was assessed by SDS-PAGE analysis.

This universal platform affords the generation of diverse functional ELP diblock copolymers with amino acid analogues selected for the incorporation based on the amber suppression strategy. In addition, we propose that the temperature could promote the self-

assembly of the photo-crosslinkable copolymers into a micelle-like structure, where the intramolecular hydrophobic interaction is greatly enhanced. After UV irradiation, the cross-linked protein micelles could retain their shapes regardless of environmental conditions. The proposal is further supported by the DLS data and TEM images where spherical structures were observed.



M: Precision Plus Dual Color Prestained Standards (Bio-Rad)
 Lane1: MRA30 Elastin-AB-UAG+ AzFRS Before induction
 Lane2: MRA30 Elastin-AB-UAG+ AzFRS (-) AzF 20hr
 Lane3: MRA30 Elastin-AB-UAG+ AzFRS (+) 1mM AzF 0hr
 Lane4: MRA30 Elastin-AB-UAG+ AzFRS (+) 1mM AzF 1hr
 Lane5: MRA30 Elastin-AB-UAG+ AzFRS (+) 1mM AzF 3hr
 Lane6: MRA30 Elastin-AB-UAG+ AzFRS (+) 1mM AzF 20hr
 Lane7-9: Purified ELP-AzF diblock copolymer

Figure 7. Expression and purification of ELP-AzF diblock copolymers. Lanes 1-6 showed the western blot analysis of the time-course expression of ELP-AzF diblock copolymers. Lanes 7-9 showed the SDS-PAGE analysis of the purified ELP-AzF diblock copolymers.

Thermo-responsive properties of ELP-AzF diblock copolymer

It was previously reported that the thermal transition properties of ELP might change greatly upon fusion to other proteins.^{41,42} To test the properties of the ELP diblock copolymers, temperature turbidity experiments were conducted on solutions of the amphiphilic diblock copolymers that had been substituted with AzF analogues. The thermal transition of the diblock copolymers was monitored by UV-visible spectrophotometry upon heating from 4 °C to 60 °C at 600 nm, that is, a wavelength that did not initiate photo-crosslinking. These temperature-dependent turbidity profiles indicated that thermally-driven self-assembly of the diblock copolymers occurred at temperatures higher than those observed for the single block elastin-AzF polymer (elastin-(AzF)₁₂), which displayed a very sharp increase above 10 °C (Fig. 8). The thermally-driven transition, as monitored by turbidity results, showed that the transition of the ELP-AzF diblock started at temperatures above 20°C, with a significant increase in the scattered light intensity due to the conformational change and the self-organization of the diblock copolymers. The shift in the transition behavior of the diblock copolymers is most likely due to the presence of the glutamic acid residues in the hydrophilic domain.²¹ 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. In addition, the assembly of the ELP-AzF diblock was shown to be reversible upon cooling (Data not shown). We hypothesize that ELP-AzF diblock copolymers could transition from unimers to nanoparticles upon temperature increase as other elastin copolymers reported previously.^{6,7} This self-assembly process is reversible and the nanoparticles are disassembled into unimers when the temperature decreases. We

therefore selected 45 °C as the working temperature for the formation of structures previously shown to be nanoparticles in the subsequent experiments.

Self-assembly of ELP-AzF diblock copolymers

To characterize the self-assembly behavior of ELP-AzF copolymers, the assembly of nanoparticles was further investigated using DLS and TEM. While the optical density is useful to determine the temperature of assembly, DLS is an useful tool to verify the size and formation of the stable nanoparticles. We propose that the increase in absorbance upon heating in the UV-visible spectrometry was caused by the self-assembly of the free chains of unimers into nanoscale micellar-like architectures induced by the selective segregation of the hydrophobic domain within the amphiphilic diblock copolymers. Therefore, we performed TEM to examine the presence and morphology of the micellar aggregates derived from the ELP-AzF copolymers. The copolymer was dissolved in aqueous solution at different pH for an hour for the completion of the self-assembly. The most prominent morphology observed in the negatively stained TEM images was spherical nanoscale particles, characterized by relatively uniform diameters (Fig. 9). The size distribution and morphology of the nanoparticles was more uniform under slightly basic conditions. In addition, under high temperature conditions, additional morphologies such as higher-order globular and cylindrical aggregates of the spherical micelles were observed. These higher order aggregates are assumed to arise from the further desolvation of the hydrophilic block and coalescence of the spherical nanoparticles. The occurrence of cylindrical micelles was also observed in solutions of synthetic amphiphilic copolymer, as well as those derived from ELPs.¹⁴ The further aggregation of the micellar particles into an organized matrix, rather than a non-uniform dispersion over the grid

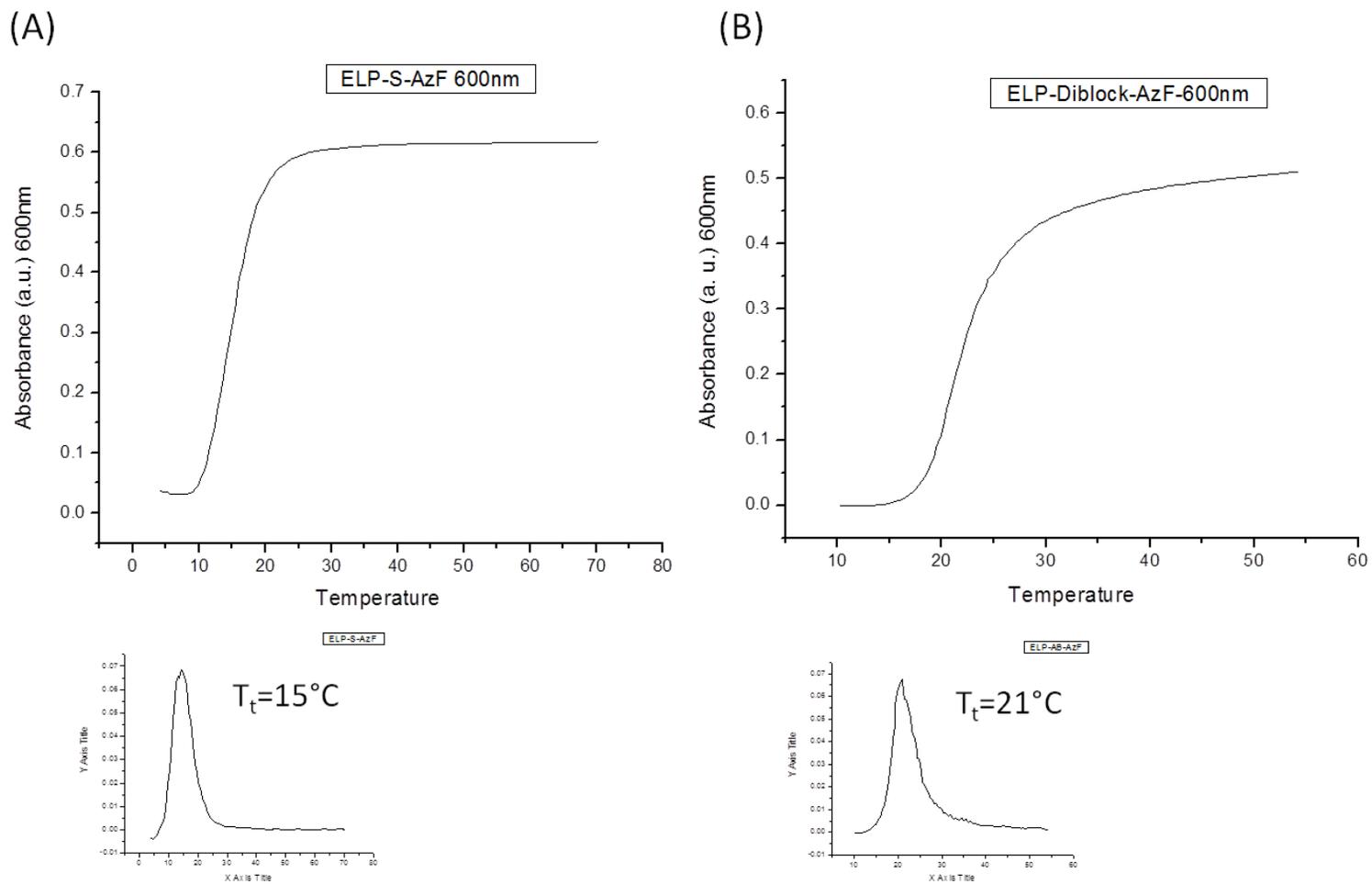


Figure 8. Temperature-dependent turbidimetry profiles of elastin-(AzF)₁₂ (A) and ELP-AzF diblock copolymers (B). The smaller plots showed the first derivatives of the turbidimetry results.

surface, could potentially be enhanced by evaporative concentration during deposition of the copolymer solution. Single-block elastin-(AzF)₁₂ was also performed as negative control for TEM studies (Fig. 10). There are no detectable nanoparticles observed and the morphology observed was unimers or irregular aggregates. Since the assembly is reversible, the nanoparticles should be disassembled into unimers when the temperature is decreased. We demonstrated the reversibility by incubating the micellar diblock solutions at 4 °C shaking overnight. The behavior of the ELPs was investigated initially by TEM to monitor the disappearance of the nanoparticles. In parallel, photo-crosslinking was performed on the ELP-AzF copolymer solution, which was incubated at higher temperature for self-assembly followed by UV-irradiation at 360nm for an hour to complete the photo-reaction. After photo-crosslinking, the ELP-AzF copolymer solution was incubated at 4 °C overnight to monitor the behavior of deformation. TEM data showed the photo-crosslinked micelles maintain their spherical shapes with radius of around 40 nm after 4 °C incubation compared to the disappearance of non-crosslinked nanoparticles (Fig. 10). Photo-crosslinking was performed at 4 °C as control and few micelles were detected by TEM (Data not shown). In addition, the morphology of the cross-linked nanoparticles was confirmed by AFM, indicating the formation of micellar-like particles (Fig. 11). However, the images obtained from the AFM tapping mode showed the flattened projection instead of the height of these nanoparticles. As the particles were pressed by the AFM tips, the measurement of the height from the particles might not be accurate and the indicated diameter, which is referred to the cross-section of these particles, could be over-estimated.

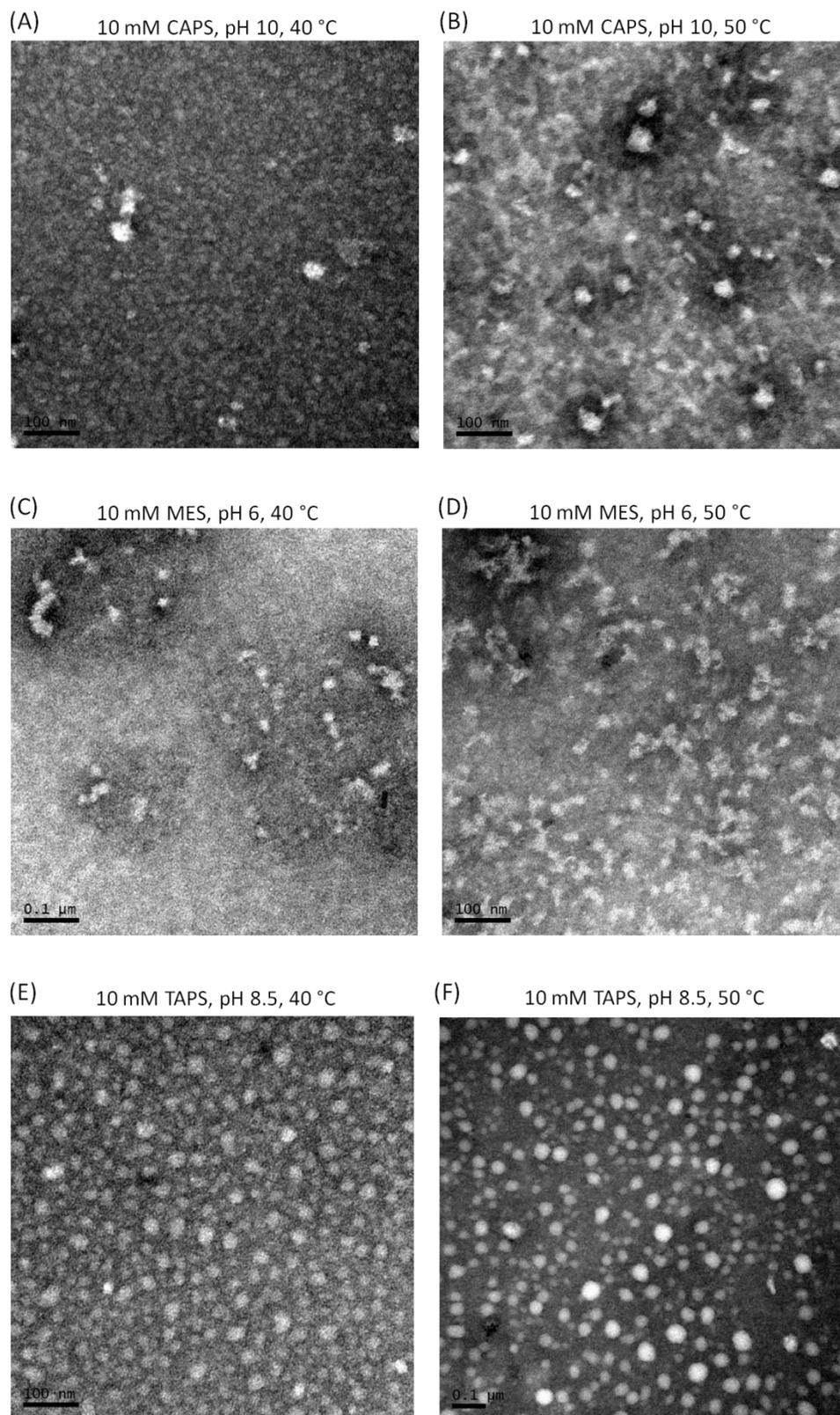


Figure 9. TEM images of ELP-AzF copolymers at various pHs and temperatures.

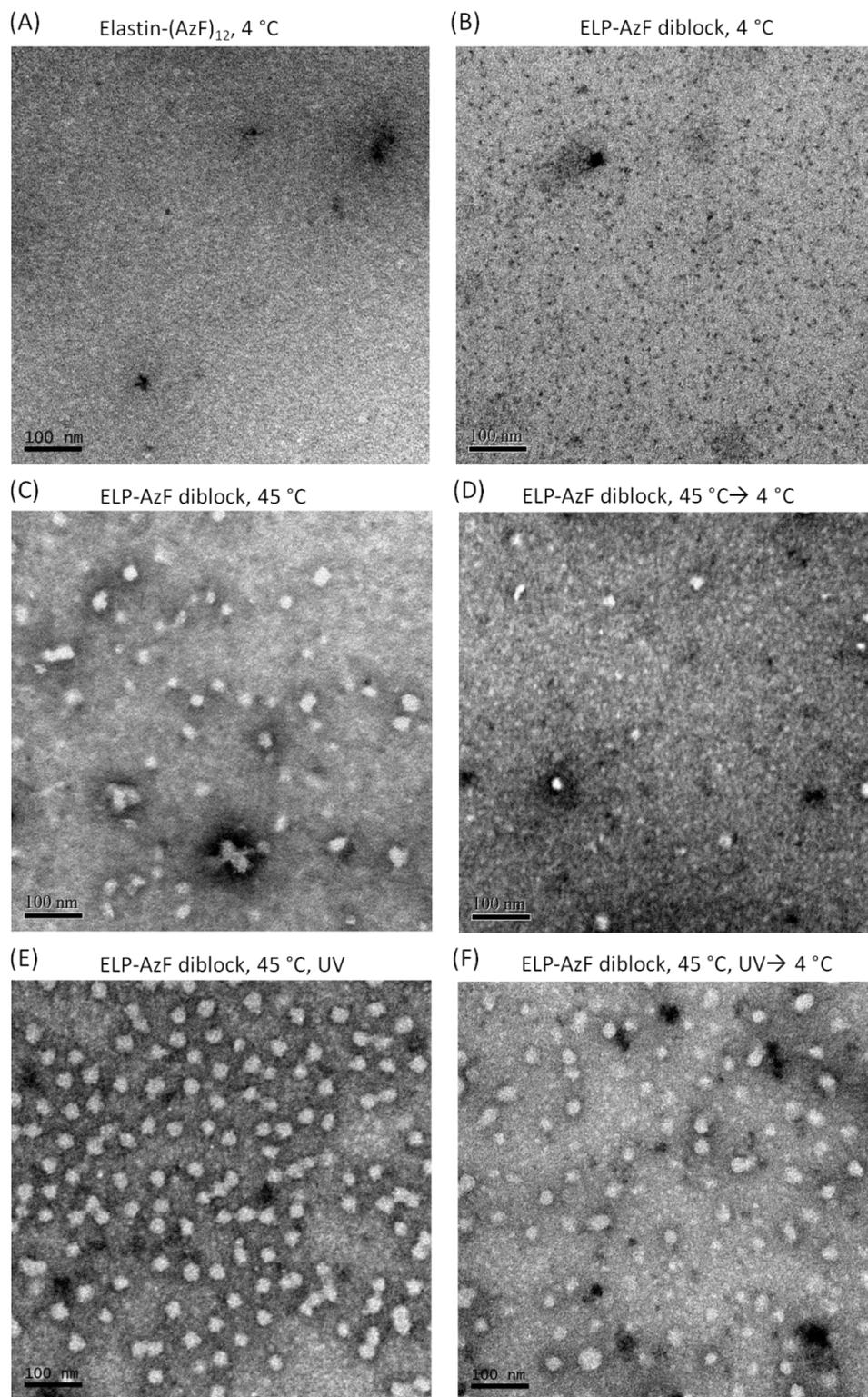


Figure 10. TEM images of the self-assemble behaviors of the ELP-AzF diblock copolymers with or without UV crosslinking (10mM TAPS buffer, pH8.5).

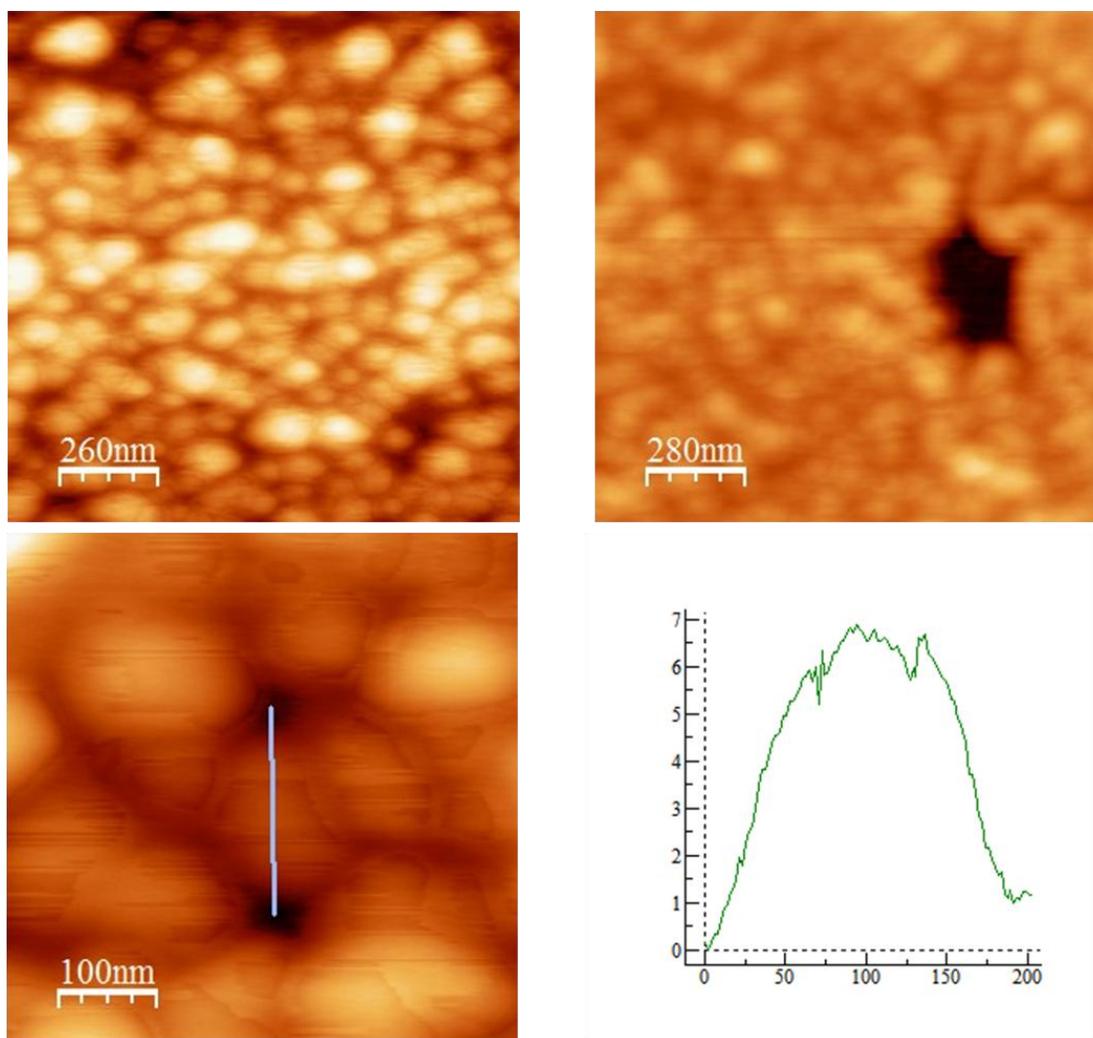


Figure 11. AFM images of the photo-crosslinked ELP-AzF diblock copolymers. The plot shows the height and the cross-section of the selected nanoparticle (nm).

We further characterized the thermal behavior and the size distribution of the nanoparticles by DLS. Initially, the ELP-AzF copolymers exhibit two phases in the solution when temperature rose from 4 °C to 45 °C (Fig. 12). The ELP-AzF copolymers at 4 °C displayed a hydrodynamic diameter (D_h) of circa 10-15 nm, which is suggestive of the presence of free chains in solution. Upon heating, the ELP-AzF copolymers transitioned from the unimers to nanoparticles with D_h in the range of 30-50 nm. Notably,

the mean particle sizes from TEM were slightly smaller than the measurements from DLS, which is probably because the deposition of the stains might be only on the surface of the particles. The morphology, shapes and consequently the sizes of particles could be also altered as a result of, for example, the dehydration during the sample preparation for TEM. Additionally, there was only one detectable population for each ELP-AzF copolymer solution at all temperatures, indicating that the self-assembly proceeded nearly to completion under these conditions. The photo-crosslinked micelles showed an improved stability toward temperature variation with the diameter around 70 nm, which is slightly larger than the formation at the working temperature. The swelling behavior could be due to the rehydration of the crosslinked hydrophobic domain upon the decrease of temperature. These data support our hypothesis that these photo-crosslinked ELP-AzF diblock nanoparticles could sustain their terminal network along with the capability of stretching-swelling responsive to outside stimuli.

We also employed the gel experiments to support the evidence for photo-crosslinking. The non-crosslinked ELP-AzFs at low (4 °C) and high temperature (45 °C) observed by the SDS-PAGE appeared at the same position that is larger than the calculated molar masses as previous studies. Photo-crosslinked ELP-AzF copolymers at 4 °C indicated the same migration as the unimer-conformation. However, photo-crosslinked ELP-AzF copolymers at 45 °C showed the aggregation at high molecular weight that could not be separated by electrophoresis (Fig. 13). In addition, we hypothesize that the photo-crosslinked ELP nanoparticles, compared to the non-crosslinked ELP copolymers, might be more resistant to the protease digestion. In our

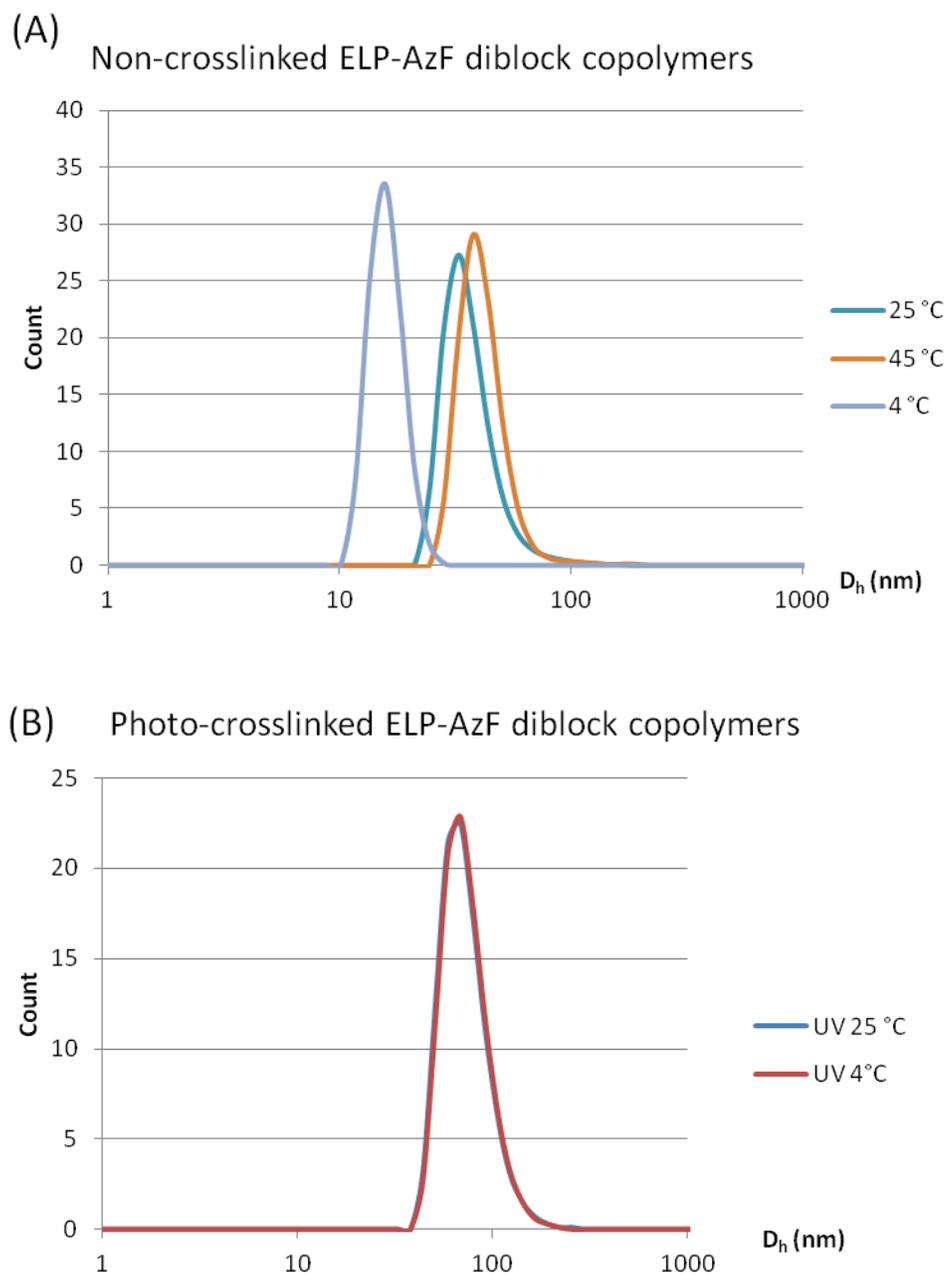


Figure 12. DLS of the ELP-AzF diblock copolymers. (A) ELP-AzF diblock copolymers at various temperatures. (B) Photo-crosslinked ELP-AzF diblock copolymers (performed at 45 °C) at various temperatures.

preliminary designs, thermolysin digestion was performed on both of the photo-crosslinked and non-crosslinked ELP-AzF copolymers. The fragments resulted from the protease digestion were analyzed by mass spectrometry. Without photo-crosslinking, the major peaks in the MALDI spectrum corresponded to the peptide fragments digested from thermolysin (Fig. 14). The azido group of the AzF might be reduced to primary amine after dithiothreitol treatment prior to the protease digestion. In contrast, once the ELP-AzF copolymers were locked in the nanoparticle structures, there were no obvious digested fragments detected in the mass spectrometer. These preliminary data suggest that the crosslinked ELP-AzF nanoparticles are more resistant to the thermolysin treatment.

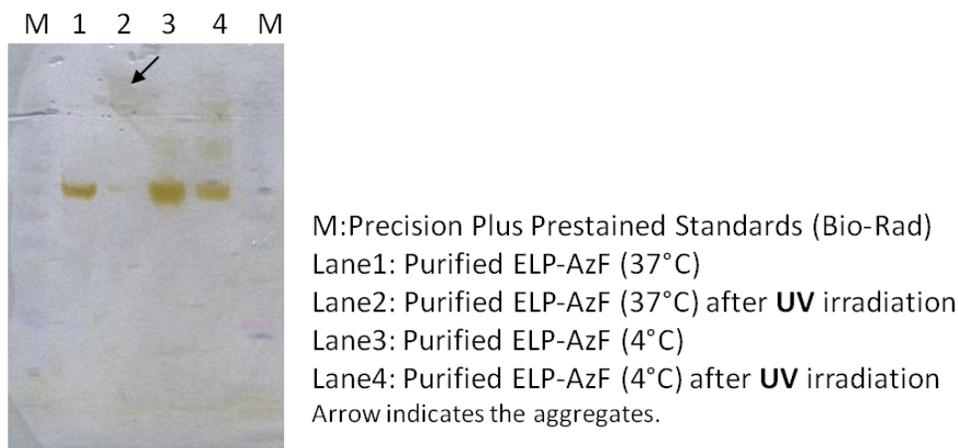


Figure 13. Gel electrophoresis of ELP-AzF diblock copolymer after UV-induced cross-linking.

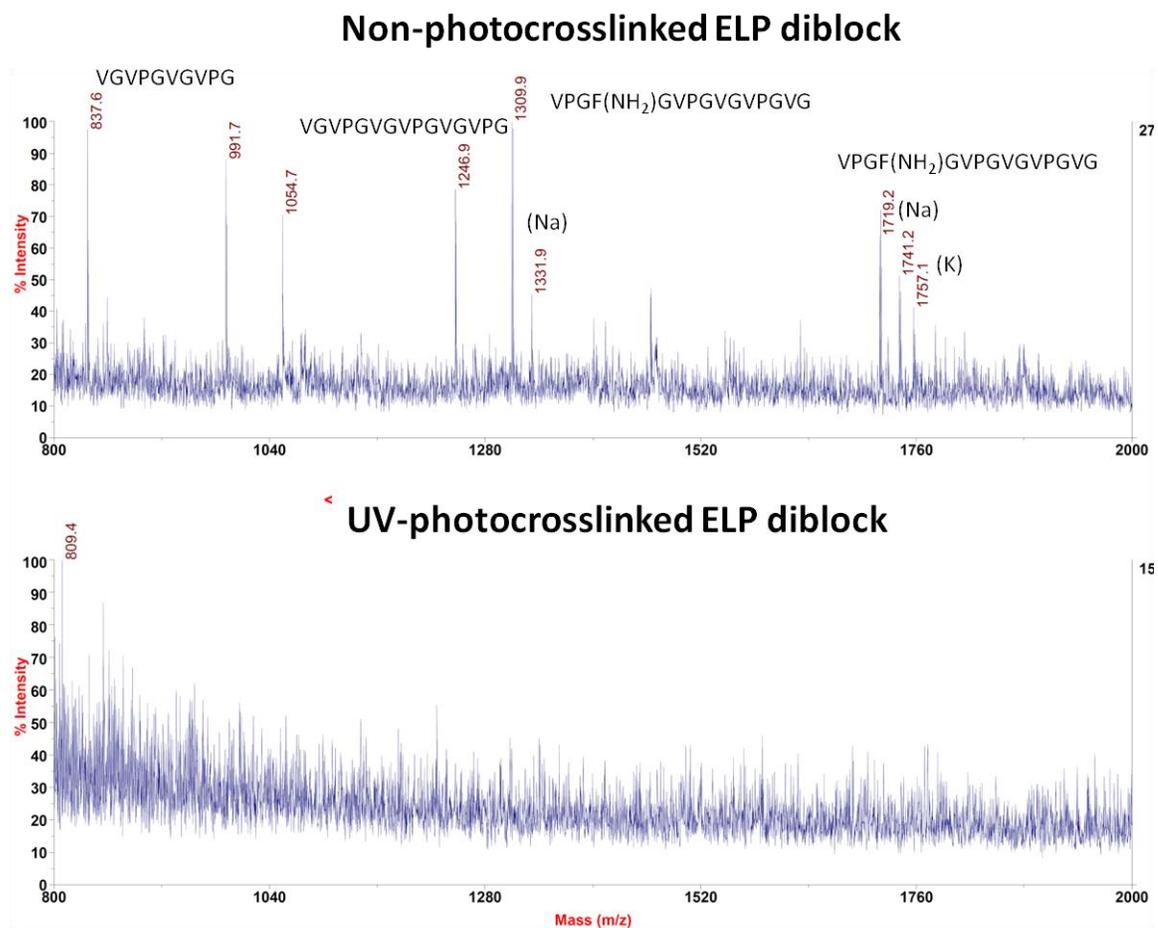


Figure 14. MALDI-TOF spectra of thermolysin digested non-crosslinked and UV-crosslinked ELP-AzF diblock copolymers.

Conclusion

Stimuli-responsive polymers have received tremendous attention in a variety of applications, such as biomedicine, diagnostic biosensors, and tissue engineering. Protein micelles based on the self-assembly of ELP amphiphilic block copolymers have significant advantages in drug delivery over conventional polymers, including their modular design, correlation between sequence and structure, potential for integration of multiple cell-responsive domains, and biocompatibility. In this study, we have demonstrated the potential of a genetically encoded polypeptide platform, in which the hydrophobic domain can be altered without the need for covalent attachment chemistry. We have successfully incorporated photo-crosslinkable amino acids into the ELP-diblock copolymers site-specifically via the amber suppression strategy. The recombinant amphiphilic diblock polypeptides can form thermally responsive micellar nanoparticles in aqueous solutions, which exhibit a spherical core-shell structure. In addition, the terminal azido group within the protein-based polymer can serve as attractive sites for the photo-initiated crosslinking under physiologically relevant conditions or further derivatization through the versatile chemistry of azides. Our results have demonstrated that amphiphilic micelles can undergo photo-crosslinking achieved through UV-irradiation, which provides structural stability to the self-assembled state even at temperatures below T_i . These photo-crosslinkable nanoparticles could facilitate the design of drug-delivering micelles that would be stable under complex physiologic conditions.

In conclusion, we have successfully constructed a new breed of ELP-based diblock copolymers featuring the unnatural amino acids for further crosslinking reaction or other down-stream applications. In addition, multiple site-specific incorporation of

unnatural amino acids using the amber suppression strategy enables incorporation of a diverse range of non-canonical amino acids with complete incorporation and uniform polypeptide composition. The controlled genetic synthesis and purification of the biopolymer that is armed by drug loading potentials, tunable thermo-responsiveness, and particle stability further present a novel platform promising for advanced biomaterial development. We envision that these novel ELPs containing non-natural amino acid analogues have advantages over previously designed protein-based materials.

Reference

- [1] W. Kim, *Front Biosci* **2013**, *18*, 289-304.
- [2] X. Hu, P. Cebe, A. S. Weiss, F. Omenetto, D. L. Kaplan, *Materials Today* **2012**, *15*, 208-215.
- [3] O. S. Rabotyagova, P. Cebe, D. L. Kaplan, *Biomacromolecules* **2011**, *12*, 269-289.
- [4] W. Kim, E. L. Chaikof, *Adv. Drug Delivery Rev.* **2010**, *62*, 1468-1478.
- [5] S. R. MacEwan, A. Chilkoti, *Biopolymers* **2010**, *94*, 60-77.
- [6] M. R. Dreher, A. J. Simnick, K. Fischer, R. J. Smith, A. Patel, M. Schmidt, A. Chilkoti, *J. Am. Chem. Soc.* **2008**, *130*, 687-694.
- [7] D. J. Callahan, W. Liu, X. Li, M. R. Dreher, W. Hassouneh, M. Kim, P. Marszalek, A. Chilkoti, *Nano Lett.* **2012**, *12*, 2165-2170.
- [8] M. S. Muthu, C. V. Rajesh, A. Mishra, S. Singh, *Nanomedicine* **2009**, *4*, 657-667.
- [9] Y. Chen, P. Youn, D. Y. Furgeson, *J. Control. Release* **2011**, *155*, 175-183.
- [10] W. Kim, J. Xiao, E. L. Chaikof, *Langmuir* **2011**, *27*, 14329-14334.
- [11] J. Andrew MacKay, M. Chen, J. R. McDaniel, W. Liu, A. J. Simnick, A. Chilkoti, *Nat. Mater.* **2009**, *8*, 993-999.
- [12] E. Blanco, C. W. Kessinger, B. D. Sumer, J. Gao, *Exp. Biol. Med.* **2009**, *234*, 123-131.
- [13] H.-C. Tsai, W.-H. Chang, C.-L. Lo, C.-H. Tsai, C.-H. Chang, T.-W. Ou, T.-C. Yen, G.-H. Hsiue, *Biomaterials* **2010**, *31*, 2293-2301.
- [14] T. A. T. Lee, A. Cooper, R. P. Apkarian, V. P. Conticello, *Adv. Mater. (Weinheim, Ger.)* **2000**, *12*, 1105-1110.
- [15] D. E. Meyer, A. Chilkoti, *Biomacromolecules* **2002**, *3*, 357-367.

- [16] N. Badi, J. F. Lutz, *Chem. Soc. Rev.* **2009**, *38*, 3383-3390.
- [17] R. E. Sallach, W. Cui, J. Wen, A. Martinez, V. P. Conticello, E. L. Chaikof, *Biomaterials* **2009**, *30*, 409-422.
- [18] W. Kim, J. Thévenot, E. Ibarboure, S. Lecommandoux, E. L. Chaikof, *Angew. Chem. Int. Ed.* **2010**, *49*, 4257-4260.
- [19] D. W. Urry, C.-H. Luan, C. M. Harris, T. M. Parker, in *Protein-Based Materials* (Eds.: D. Kaplan, K. McGrath), Birkhauser, Boston, **1997**, pp. 133-177.
- [20] D. W. Urry, D. C. Gowda, T. M. Parker, C. H. Luan, M. C. Reid, C. M. Harris, A. Pattanaik, R. D. Harris, *Biopolymers* **1992**, *32*, 1243-1250.
- [21] D. W. Urry, C. H. Luan, T. M. Parker, D. C. Gowda, K. U. Prasad, M. C. Reid, A. Safavy, *J. Am. Chem. Soc.* **1991**, *113*, 4346-4348.
- [22] V. P. Conticello, H. E. Carpenter Desai, in *Polymer Science: A Comprehensive Reference Vol. 9* (Eds.: K. Matyjaszewski, M. Möller), Elsevier, **2012**, pp. 71-103.
- [23] W. Hassouneh, K. Fischer, S. R. MacEwan, R. Branscheid, C. L. Fu, R. Liu, M. Schmidt, A. Chilkoti, *Biomacromolecules* **2012**, *13*, 1598-1605.
- [24] W. Kim, C. Brady, E. L. Chaikof, *Acta Biomater.* **2012**, *8*, 2476-2482.
- [25] G. Sun, P.-Y. Hsueh, S. M. Janib, S. Hamm-Alvarez, J. Andrew MacKay, *J. Control. Release* **2011**, *155*, 218-226.
- [26] A. J. Simnick, C. A. Valencia, R. Liu, A. Chilkoti, *ACS Nano* **2010**, *4*, 2217-2227.
- [27] S. R. MacEwan, A. Chilkoti, *Nano Lett.* **2012**, *12*, 3322-3328.
- [28] A. J. Simnick, M. Amiram, W. Liu, G. Hanna, M. W. Dewhirst, C. D. Kontos, A. Chilkoti, *J. Control. Release* **2011**, *155*, 144-151.

- [29] A. Guo, G. Liu, J. Tao, *Macromolecules* **1996**, *29*, 2487-2493.
- [30] K. B. Thurmond, T. Kowalewski, K. L. Wooley, *J. Am. Chem. Soc.* **1996**, *118*, 7239-7240.
- [31] S. C. Payne, M. Patterson, V. P. Conticello, in *Protein Engineer Handbook, Vol. 2* (Eds.: S. Lutz, U. T. Bornscheuer), Wiley-VCH, **2009**, pp. 915-938.
- [32] G. T. Hermanson, *Bioconjugate Techniques, 2nd Edition* **2008**, 1-1202.
- [33] J. Jiang, B. Qi, M. Lepage, Y. Zhao, *Macromolecules* **2007**, *40*, 790-792.
- [34] I. S. Carrico, S. A. Maskarinec, S. C. Heilshorn, M. L. Mock, J. C. Liu, P. J. Nowatzki, C. Franck, G. Ravichandran, D. A. Tirrell, *J. Am. Chem. Soc.* **2007**, *129*, 4874-4875.
- [35] S. Wang, C. Wong Po Foo, A. Warriar, M. M. Poo, S. C. Heilshorn, X. Zhang, *Biomed. Microdevices* **2009**, *11*, 1127-1134.
- [36] P. J. Nowatzki, C. Franck, S. A. Maskarinec, G. Ravichandran, D. A. Tirrell, *Macromolecules* **2008**, *41*, 1839-1845.
- [37] J. Raphel, A. Parisi-Amon, S. C. Heilshorn, *J. Mater. Chem.* **2012**, *22*, 19429-19437.
- [38] M. A. Patterson, Emory University **2011**.
- [39] S. Zhang, M. Ryden-Aulin, L. A. Kirsebom, L. A. Isaksson, *J. Mol. Biol.* **1994**, *242*, 614-618.
- [40] K. Trabbic-Carlson, L. A. Setton, A. Chilkoti, *Biomacromolecules* **2003**, *4*, 572-580.
- [41] X. X. Xia, Q. Xu, X. Hu, G. Qin, D. L. Kaplan, *Biomacromolecules* **2011**, *12*, 3844-3850.

- [42] M. Dai, J. Haghpanah, N. Singh, E. W. Roth, A. Liang, R. S. Tu, J. K. Montclare, *Biomacromolecules* **2011**, *12*, 4240-4246.

Chapter 5

Conclusion

Conclusion

Biosynthesis of these sequence-repetitive polypeptides derived from native structural proteins has been the subject of significant attention for its potential as a biomaterial in tissue-engineering and biotechnological applications.¹⁻⁴ However, chemical functionalities available via protein translation is modest and is restricted by the limited library of amino acids specified by the genetic code, particularly in comparison to organic polymers derived from chemical synthesis. The introduction of new building blocks displaying unique functional groups with novel patterns of reactivity significantly expands the scope and impact of protein material engineering.⁵⁻⁷ The development of methods that allow multi-site specific incorporation of unnatural amino acids represents a significant advance in biosynthetic technology, especially in engineering protein-based polymers consisting of monomers with expanded chemical diversity. Most approaches have been targeted at enhancing the ability of protein biosynthetic machinery to utilize amino acid analogues beyond the twenty natural amino acids.⁶ In this work, we have successfully demonstrated an *E. coli* bacterial expression system for the incorporation of non-canonical amino acids at multiple pre-defined positions within the polypeptide chains. Furthermore, we have also illustrated the application of this system in the original design of novel elastin-like protein (ELP) materials.

Schultz and coworkers have reported the co-translational incorporation of a new set of chemically diverse tyrosine derivatives into single amber sites encoded within recombinant polypeptides.⁸ In addition, several research laboratories have described mutant versions of pyrrolysyl-tRNA synthetases (PylRS) that enable the insertion of functionally modified amino acid derivatives using a similar approach.⁹⁻¹¹ Due to the

variety of currently available orthogonal unnatural-aminoacyl-tRNA synthetases (UaaRS)/tRNA_{CUA} pairs, the amber suppression strategy is targeted as defined sites for the multi-site specific incorporation of selected unnatural amino acids in this study. However, since there will be competition between amber codons and the endogenous termination machinery, this scheme might be challenging for the multiple-site selective incorporation of analogues. Therefore, we demonstrate a strategy for increased amber suppression that employs a genetically modified *E. coli* strain, MRA 30, with an attenuated RF1 activity¹² as our bacterial host system. We hypothesized that the increased readthrough of the stopcodon UAG in the modified expression strain should permit multiple amber suppression events during protein translation. To verify our hypothesis, we built up the system with the following components. First, we selected PylRS pairs from *Methanosarcina barkeri fusaro* as our model orthogonal system because of the convenience of its natural orthogonality in *E. coli* and the fact that the native tRNA^{pyl}_{CUA} decodes the amber codon without any modification.¹³ Mutant PylRS pairs that could activate several *N*- ϵ -alkoxycarbonyl-substituted lysine derivatives for the co-translational insertion with high efficiency⁹ were engineered to demonstrate the feasibility of multiple amber suppression within the polypeptide chain. In addition, a robust fluorescent reporter protein was modified as our model protein to assess the suppression efficiency in the proposed system. A FACS screen based on the suppression of a triple non-sense insertion mutant of sfGFP¹⁴ was utilized to monitor the suppression efficiency as the fluorescence was proportional to the expression of the full-length protein. The experimental results have provided evidence that *E. coli* strains with an attenuated RF1 activity are competent for the multi-site insertion of non-canonical amino acids at amber codon sites encoded

within the corresponding genetic template. At this point, we believe that this system is useful for the selected incorporation of amino acid analogues and the sfGFP-(UAG)_n template can therefore serve as a robust readout for designing a high-throughput screening strategy for the direct evolution of particular aaRSs.

Next, we brought this system into the design of novel protein-based materials with extended chemical functionalities. Elastin-mimetic peptides, [(Val-Pro-Gly-Val-Gly)₂(Val-Pro-Gly-Xaa-Gly)(Val-Pro-Gly-Val-Gly)₂]_n, where amber codons (Xaa) are periodically located at the guest position, were generated as genetic templates to evaluate the feasibility of the preparation of protein based materials, which are usually derived from sequence-repetitive polypeptides. We have not only characterized the expression level of these elastin derivatives but also ascertained the site-specific localization of the amino acid analogues within the polypeptide chain via mass spectrometry analysis. Our data suggest that the attenuation of RF1 activity can significantly enhance the incorporation of non-canonical amino acids through the decreased competition with aminoacylated suppressor tRNAs for the recognition of the amber codon sites encoded within the translational reading frame of the recombinant polypeptide. The efficacy of the process depends strongly on the identity of the orthogonal aaRS/suppressor tRNA pair and the local sequence context.^{15,16} Significant levels of the recombinant protein expression were detected within the sequence-repetitive proteins encoding up to 22 amber suppression sites. While the process remains to be optimized with respect to the individual analogues, these results encourage that sequence-repetitive polypeptides derived from the native protein materials can be prepared with high levels of incorporation of non-canonical amino acids with the selectivity that overcomes those of

the residue-specific approaches. These substitutions that occur within a uniform and controllable sequence enhance the potential for coupling the chemical properties of the variant side-chains to the materials properties of protein polymers.

In an attempt to examine the generality of this approach, the multi-site substitution was attempted using an alternative amber suppression system based on an orthogonal *M. jannaschii* tyrosyl-tRNA synthetase/MjtRNA_{CUA} (MjTyrRS) pair.⁸ Schultz and coworkers have developed methods for the mutagenesis and selection of mutant tyrosyl-tRNA synthetases that are capable of the activation for a chemically diverse range of substituted aromatic amino acids in *E. coli* host strains. In addition, the wild-type *M. jannaschii* tRNA(MjtRNA_{CUA}) has been extensively modified to enhance its orthogonality in *E. coli* and to serve as a more effective suppressor of the amber codon.^{17,18} This orthogonal system has been employed to incorporate non-canonical amino acids at a single defined site within recombinant polypeptides in a reasonably high yield. In this study, the Bpa and AzF residues are of particular interest for the multi-site substitution due to their potential as photo-crosslinkers.¹⁹⁻²¹ Furthermore, the Bpa and AzF residues provide photo-crosslinkable sites that may facilitate the processing of the corresponding elastin-mimetic protein polymers into elastin networks under environmentally benign conditions.⁵ We, therefore, engineered the orthogonal pairs with respect to the Bpa and AzF analogues on the basis of optimal MjtRNA_{CUA}s. The sfGFP variant, with a triple non-sense insertion mutant, was employed to evaluate the suppression activity of these orthogonal pairs. Of the synthetase variants we tested, the MjtRNA_{CUA}-Nap1 could be used to decode amber stop codons with higher efficiency. Adapting these optimized orthogonal pairs, elastin-mimetic polypeptides encoding amber codons specified for Bpa

or AzF residues could be expressed and isolated with significant levels regarding the density of termination codons within the polypeptide chains. In the last chapter, we demonstrated one of the applications of our system to the design of photo-crosslinkable ELP amphiphilic diblock copolymers. The AzF residues, as photo-crosslinkers, were selectively incorporated into the hydrophobic domain, and the expression of the recombinant ELP copolymers could be detected and isolated. The ELP diblock copolymers would reversibly transit between the states of unimers and micellar nanoparticles. In addition, UV radiation promotes the photo-crosslinking of the ELP copolymers and processes the corresponding copolymers into lock-up elastin networks. In this case, our crosslinked ELP copolymers retained their stable terminal structures as nanoparticles, along with the swelling-stretching behaviors from elastin-mimetic polypeptides. Therefore, incorporation of other cell-instructive domains in the photo-crosslinkable ELPs could be beneficial in further applications for targeting drug delivery.²²

Overall, the biosynthetic approach (Fig. 1) described in this study enables the preparation of sequence-repetitive polypeptides, in which multiple non-canonical amino acids could be incorporated at pre-defined positions in the polypeptide sequence. The multi-site suppression method leverages the utility of these systems for the preparation of substituted protein polymers, in which the material properties can be enhanced through the incorporation of novel chemical functionality at precisely defined positions within the sequence. In addition, the repetitive sequences derived from many native protein-based materials might include multiple sites of post-translational modifications that are critical for appropriate biological function, for example, *N*- ϵ -acetyl-lysine in histone proteins²³,

trans-hydroxyproline in collagens²⁴, or DOPA residues in mussel adhesive proteins.²⁵ In conclusion, the multi-site suppression techniques represent an attractive approach to the synthesis of these functional materials which are challenging to obtain from conventional *in vitro* chemical or *in vivo* enzymatic modifications of the native protein sequences.

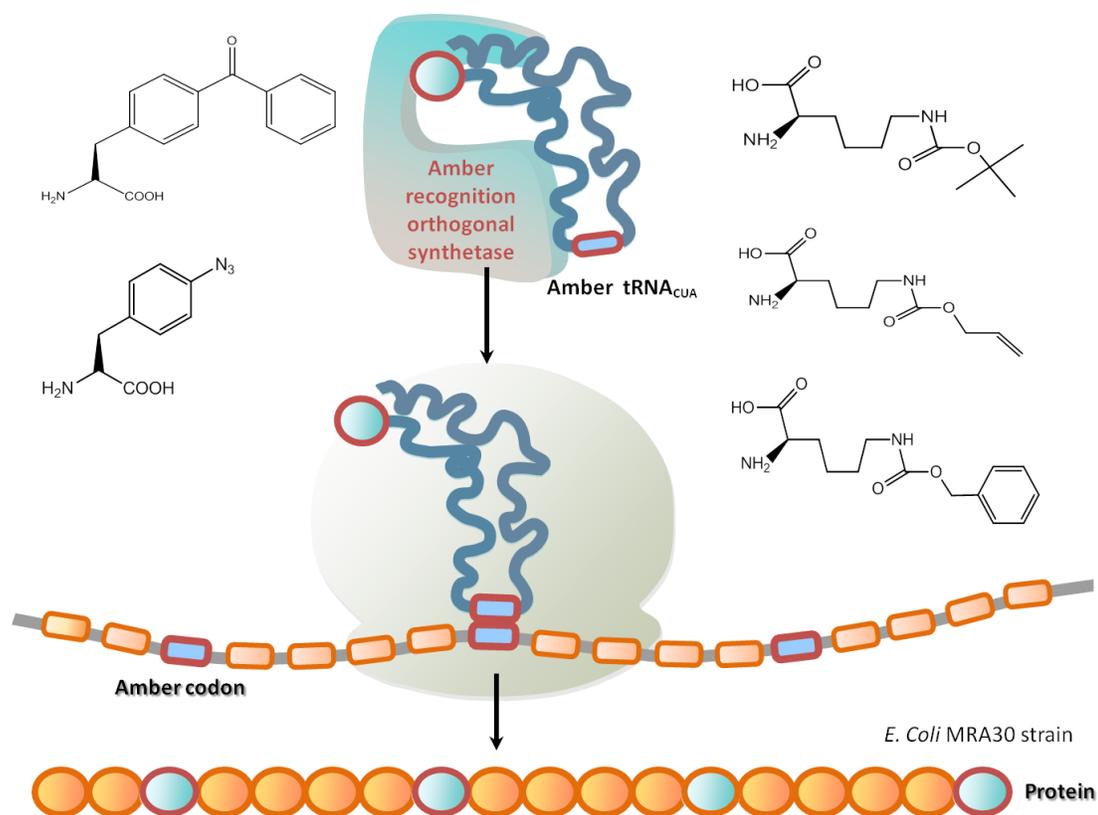


Figure 1. Multi-site specific incorporation of non-canonical amino acids in *E. coli*. A genetic modified *E. coli* MRA30 strain with attenuated RF1 activity (*prfA1*) is used as an expression host. Pyrrolysine analogues and phenylalanine analogues are selected for the demonstration of multi-site specific incorporation into polypeptide chains.

Reference

- [1] S. R. MacEwan, A. Chilkoti, *Biopolymers* **2010**, *94*, 60-77.
- [2] W. Kim, E. L. Chaikof, *Adv. Drug Delivery Rev.* **2010**, *62*, 1468-1478.
- [3] V. P. Conticello, H. E. Carpenter Desai, in *Polymer Science: A Comprehensive Reference Vol. 9* (Eds.: K. Matyjaszewski , M. Möller), Elsevier, **2012**, pp. 71-103.
- [4] N. Srinivasan, S. Kumar, *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **2012**, *4*, 204-218.
- [5] K. Zhang, M. R. Diehl, D. A. Tirrell, *J. Am. Chem. Soc.* **2005**, *127*, 10136-10137.
- [6] J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, D. A. Tirrell, *Curr. Opin. Chem. Biol.* **2010**, *14*, 774-780.
- [7] M. A. Patterson, Emory University **2011**.
- [8] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413-444.
- [9] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187-1197.
- [10] Y.-M. Li, M.-Y. Yang, Y.-C. Huang, Y.-T. Li, P. R. Chen, L. Liu, *ACS Chem. Biol.* **2012**, *7*, 1015-1022.
- [11] T. Kobayashi, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2009**, *385*, 1352-1360.
- [12] S. Zhang, M. Ryden-Aulin, L. A. Kirsebom, L. A. Isaksson, *J. Mol. Biol.* **1994**, *242*, 614-618.
- [13] O. Namy, Y. Zhou, S. Gundllapalli, C. R. Polycarpo, A. Denise, J.-P. Rousset, D. Söll, A. Ambrogelly, *FEBS Lett.* **2007**, *581*, 5282-5288.

- [14] J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **2006**, *24*, 79-88.
- [15] L. Bossi, *J. Mol. Biol.* **1983**, *164*, 73-87.
- [16] S. Zhang, M. Ryden-Aulin, L. A. Isaksson, *FEBS Lett.* **1999**, *455*, 355-358.
- [17] T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, *395*, 361-374.
- [18] J. Guo, C. E. Melancon, 3rd, H. S. Lee, D. Groff, P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9148-9151.
- [19] J. C. Kauer, S. Erickson-Viitanen, H. R. Wolfe, Jr., W. F. DeGrado, *J. Biol. Chem.* **1986**, *261*, 10695-10700.
- [20] J. W. Chin, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020-11024.
- [21] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026-9027.
- [22] J. Raphel, A. Parisi-Amon, S. C. Heilshorn, *J. Mater. Chem.* **2012**, *22*, 19429-19437.
- [23] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232-234.
- [24] D. D. Buechter, D. N. Paolella, B. S. Leslie, M. S. Brown, K. A. Mehos, E. A. Gruskin, *J. Biol. Chem.* **2003**, *278*, 645-650.
- [25] S. Lim, K. R. Kim, Y. S. Choi, D. K. Kim, D. Hwang, H. J. Cha, *Biotechnol. Prog.* **2011**, *27*, 1390-1396.

Appendix

Sequence of Interest

Appendix 1. Sequencing primers utilized in the study.

Sequence primers	Sequence (5'→3')
pQE-F	CCCGAAAAGTGCCACCTG
pQE-R	GGTCATTACTGGAGTCTTG
pSU81-F	GCTGACGCACCGGTGCAGCC
pSU81-R	GCTTCTGTTTCTATCAGCTG
HtRNA-F	GTCGGTGAACGCTCTCCTGAGTAGG
HtRNA-R	CGCCACCACTGATTTGAGCGTCAG
5-Pro	CACATCAGCAGGACGCACTGAC
3-Pro	CGCTCGCCGCAGCCGAAC
M13 (-20) Forward	TGTAAAACGACGGCCAGT
M13 Reverse	CAGGAAACAGCTATGAC
SP6	GAT TTAGGTGACACTATAG
T7	TAATACGACTCACTATAGGG

Appendix 2. Sequences of the MbPylRS.

>Optimized DNA sequence encoding MbPylRS from DNA2.0

```
GGTACCATGGACAAGAAACCTTTGGATGTGTTGATTTCCGGCTACTGGTTTGTGGATGAGCCGTACCGGCAC
GCTGCATAAGATTAAGCACTACGAGGTTTCGCGTTCCAAAATCTACATCGAGATGGCCTGTGGCGACCATT
TGGTCGTTAACAATAGCCGTAGCTGTCTGACCGCACGCGCATTCCGTCATCACAAGTACCGCAAAACCTGT
AAACGTTGTCGTGTCAGCGATGAAGATATCAACAACCTTTCTGACGCGCAGCACGGAGGGTAAAACCAGCGT
CAAGGTTAAAGTTGTCTCCGCACCGAAAGTGAAAAAGGCGATGCCGAAAAGCGTGAGCCGCGCACCGAAGC
CGCTGGAGAATCCGGTGAGCGCGAAGGCCTCTACTGACACCTCTCGTAGCGTTCCGTCCCCAGCCAAATCC
ACCCCGAACAGCCCAGTGCCGACGAGCGCGCCAGCCCCGAGCTTGACCCGTTCCCAACTGGATCGTGTTGA
GGCGCTGCTGAGCCCTGAGGATAAGATTTCTCTGAACATCGCTAAACCGTTTCGCGAGCTGGAAAGCGAAT
TGGTGACGCGTCGCAAGAATGATTTCCAGCGTCTGTATACCAATGACCGCGAGGACTATCTGGGCAAACCTG
GAGCGTGACATCACGAAGTTCTTTGTGGATCGTGATTTCTGGAGATTAAGAGCCCGATTCTGATTCCTGC
GGAATACGTCGAGCGCATGGGTATCAATAACGATACCGAACTGAGCAAACAAATCTTCCGTGTCGATAAGA
ATCTGTGCCTGCGTCCGATGTTGGCACCGACCCTGTACAACCTACCTGCGCAAGCTGGACCGCATTCTGCCT
GACCCGATTAAGATTTTCGAGGTTGGCCCGTGCTATCGCAAAGAAAGCGACGGTAAAGAGCACCTGGAAGA
GTTTACCATGGTGAACCTTTTGCCAGATGGGTTCCGGTTGCACGCGTGAAAATCTGGAAAGCTTGATCAAAG
AATTTCTGGACTACCTGGAAATTGACTTTGAAATCGTGGGCGATAGCTGCATGGTGTATGGTGACACTCTG
GACATCATGCACGGCGATTTGGAAGTGTCTAGCGCGGTCGTTGGTCCGGTTCCGCTGGACCGTGAGTGGGG
TATCGACAAACCGTGGATTGGTGCGGGTTTTGGTCTGGAGCGTCTGTTGAAAGTTATGCATGGCTTCAAGA
ACATCAAACGCGCGAGCCGTAGCGAGAGCTACTATAACGGTATTTTCGACCAATCTGTAATCTAGAAA
```

>Amino acid sequence of the MbPylRS

```
GTM1DKKPLDVLISATGLWMSRTGTLHKIKHYEVSRSKIYIEMACGDHLVNNRSRSCRTARAFRHH
KYRKTCKRCRVSDDEDINNFLTRSTEGKTSVKVKVVSAPKVKKAMPKSVSRAPKPLENPVSAKAST
DTSRSVPSPAKSTPNSPVPTSAPAPSLTRSQLDRVEALLSPEDKISLNIKPFRELESELVTRRK
NDFQRLYTNDREDYLGKLERDITKFFVDRDFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDK
NLCLRPLAPTLYNYLRKLDRIILPDPIKIFEVGPCYRKESDGKEHLEEFMTMVNFCQMGSGCTREN
LES LIKEFLDYLEIDFEIVGDSCMVYGDITLDIMHGDELELSSAVVGPVPLDREWGIDKPWIGAGFG
LERLLKVMHGFKNIKRASRSESYNGISTNL*SRX
```

Appendix 3. Sequences of the MjTyrRS.

>Optimized DNA sequence encoding MjTyrRS from DNA2.0.

GGTACCATGGACGAGTTCGAAATGATCAAAAGAAACACCAGCGAAATCATCTCCGAAGAAGAGCT
 GCGTGAGGTTCTGAAGAAGGATGAGAAGAGCGCGTATATTGGCTTCGAGCCAAGCGGTAAAATCC
 ACCTGGGTCACTACCTGCAAATCAAGAAGATGATTGACCTGCAAAATGCCGGCTTCGACATCATT
 ATTTCTGCTGGCTGACCTGCATGCGTACCTGAATCAGAAAGGTGAATTGGACGAGATTTCGTAAGAT
 TGGTGACTATAACAAAAAGGTGTTTCGAGGCGATGGGTCTGAAGGCCAAAATACGTGTATGGCTCTG
 AGTTTCAGCTGGACAAAGATTACACCCTGAATGTTTATCGTCTGGCCCTGAAAACCACTCTGAAG
 CGTGCTCGCCGCTCCATGGAGTTGATTGCCCGTGAAGATGAAAACCCGAAAGTCGCCGAAGTCAT
 TTACCCGATTATGCAAGTTAACGATATCCACTACCTGGGCGTGGATGTGGCGGTTCGGTGGTATGG
 AGCAGCGCAAATCCATATGCTGGCACGCGAGCTGCTGCCGAAAAGGTCGTGTGCATCCACAAC
 CCGGTACTGACGGGTCTGGACGGTGAGGGTAAGATGAGCTCGAGCAAGGGCAACTTCATTGCGGT
 GGACGATAGCCCGGAAGAGATCCGCGCGAAAATCAAGAAAGCGTATTGTCCGGCAGGCCTTGTCG
 AAGGTAATCCGATTATGGAGATTGCAAAATACTTTCTGGAGTACCCGTTGACCATCAAGCGTCCG
 GAGAAATTTGGCGGTGATCTGACGGTTAATAGCTATGAAGAGTTGGAGAGCCTGTTTAAGAACAA
 AGAGCTGCATCCTATGGATTTGAAGAATGCAGTTGCGGAAGAACTGATCAAAATCTTGAACCGA
 TTCGTAAACGTCTGTAATCTAGAAA

>Amino acid sequence of the MjTyrRS

GTMDEFEMIKRNTSEIIEEELREVLKKDEKSAYIGFEPGKIHLGHYLQIKKMIDLQNAGFDII
 ILLADLHAYLNQKGELEIRKIGDYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTLK
 RARRSMELIAREDENPKVAEVIYPIMQVNDIHYLGVDVAVGGMEQRKIHMLARELLPKKVVCIHN
 PVLTLGLDGEKMSSSKGNFIAVDDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP
 EKFGGDLTVNSYEELESFLKKNELHPMDLKNAVAEELIKILEPIRKRL*SR

Appendix 4. Sequences of the wild-type sfGFP-CCC.

>Optimized DNA sequence encoding sfGFP-CCC from DNA2.0.

GAATTCATTAAAGAGGAGAAATTAACCATGAGCAAGGGTGAAGAGCTGTTACGGGTGTTGTGCCCAT
 TCTGGTCGAACTGGACGGTGACGTTAATGGCCATAAGTTTAGCGTGGTGAAGGTGAGGGTGACG
 CCACCAATGGTAAGTTGACCTTGAAGTTCATTTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACG
 CTGGTTACCACTCTGACGTATGGTGTGCAGTGTTCAGCCGTTACCCCGACCACATGAAGCGCCACGA
 TTTCTTCAAAAGCGCGATGCCCGAGGGCTATGTCCAGGAACGCACCATCAGCTTCAAAGATGACGGCA
 CCTACAAAACGCGTGCTGAGGTGAAGTTTGAAGGTGATACCCTGGTCAATCGCATTGAACTGAAAGGT
 ATTGACTTCAAAGAGGACGGCAATATCCTGGGTCATAAACTGGAGTACAACCTTAAACAGCCATAACGT
 TTACATCACGGCCGACAAACAAAAGAACGGCATCAAGGCAGAACTTTAAGATCCGTCACAATGTTGAGG
 ATGGTAGCGTCCAACCTGGCAGATCACTATCAGCAGAACACTCCCATTGGCGATGGCCCCGTGCTGCTG
 CCCGATAACCACTACCTGTCGACGCAATCTGTTCTGAGCAAAGATCCCAATGAGAAACGTGACCACAT
 GGTCTGCTGGAGTTTGTGACCGCAGCGGGCATTACCCATGGTATGGATGAATTGTACAAAGGATCCT
GATAAGCTT

>Amino acid sequence of the sfGFP-CCC.

EFIKKEEKLTMSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTCLKFICTTGKLPVP
 WPTLVTTLTLYGVQCFSTRYPDHMKRHDFKSSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVN
 RIELKGIDFKEDGNILGHKLEYNFNShNVYITADKQNGIKANFKIRHNVEDGQSVQLADHYQQNT
 PIGDGPVLLPDNHYLSTQSVLSKDPNEKRDMVLLLEFVTAAGITHGMDELYKGS**AX

Appendix 5. Sequences of the wild-type sfGFP-(UAG)₄.

>Optimized DNA sequence encoding sfGFP(UAG)₄ from DNA2.0.

GAATTCATTAAAGAGGAGAAATTAACCATGAGCTAGGGCGAAGAATTATTCACCGGCGTCGTACC
 AATCCTGGTTGAACTGGATGGTGAATGGTCACAAGTTTAGCGTTCGCGGTGAAGGCGAGG
 GTGACGCGACCAATGGTAAATTGACCCTGAAGTTCATTTGTACGACGGGCAAACCTGCCGGTTCCG
 TGGCCGACCCTGGTTACCACGCTGACCTACGGTGTGCAATGCTTTAGCCGTTACCCGGACCACAT
 GAAACGCCACGACTTCTTTAAGTCCGCGATGCCGGAAGTTATGTGCAAGAGCGTACTATCAGCT
 TTAAAGATGACGGTACCTATAAGACCCGTGCCGAGGTTAAGTTTGAGGGTGATACGCTGGTCAAT
 CGTATTGAGTTGAAAGGCATCGACTTCAAAGAGGACGGTAACATCCTGGGCCACAACTGGAGTA
 CAATTTCAACAGCCACAACGTGTATATCACTGCCGATAAACAGTAGAAGAACGGCATTAAAGCAA
 ATTTCAAGATCCGCCATAACGTGAGTAGGACGGTCTGTCCAGCTGGCTGACCACTACCAGCAG
 AACACCCCGATTGGTGACGGTCCTGTGCTGTAGCTGCCGATAACCATTATCTGAGCACGCAAAG
 CGTCCTGTGCGAAAGATCCGAATGAAAAGCGTGACCATATGGTGCTGTTGGAGTTCGTTACCGCAG
 CGGGCATTACGCATGGCATGGATGAACTGTACAAGGGATCCTAATAAGCTT

> Amino acid sequence of the sfGFP(UAG)₄.

EFIKKEEKL^{MS}*GEELFTGVVPI^LVELDGDVNGHKFSVRGEGEGDATNGKLT^LKFICTTGKLPVP
 WPTLVTTLT^YGVQCF^SRYPDHMKRHDF^FKSAMPEGYVQERTIS^FKDDGTYKTRAEVKFEGDTLVN
 RIELKGIDFKEDGNILGHKLEYNFNSHN^VYITADKQ^{*}KNGIKANFKIRHNVE^{*}DGSVQLADHYQQ
 NTPIGDGPVL^{*}LPDNHYLSTQSVLSKDPNEKRDHMLLEFVTAAGITHGMDELYKGS**AX