PROTEIN MODIFICATION THROUGH *IN VIVO* INCORPORATION OF NONCANONICAL AMINO ACIDS

Thesis by

Marissa Lee Mock

In Partial Fulfillment of the Requirements for the

degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2005

(Defended October 21, 2005)

ACKNOWLEDGEMENTS

I wish to thank my advisor, Professor David Tirrell, for his advice and support. I especially appreciate the laboratory atmosphere of intellectual independence with friendly collaboration that he fostered. I also wish to thank every member of the Tirrell Lab, past and present, who made my time here such a pleasure. In particular, this research could not have been accomplished without the support, both intellectual and personal, of Sarah Heilshorn, David Flanagan, Paul Nowatzki, Julie Liu, Kimberly Beatty, Kristi Kiick, Isaac Carrico, Kathy DiZio, and Rebecca Connor.

I thank my committee, Professors Dennis Dougherty, Carl Parker, and Christina Smolke for their advice and assistance.

I cannot adequately convey here how grateful I am to Sarah Heilshorn and Andrew Spakowitz, my roommates for four years, for simply everything.

I earnestly thank Lars Cremean for being there for me.

Finally, my deepest gratitude goes to my family, especially my parents, who never let me imagine that there was anything I couldn't do if I tried. They were willing to sacrifice to give me everything, and any credit for an accomplishment of mine is entirely theirs.

ABSTRACT

Traditional techniques of polymer synthesis produce macromolecules with statistical distributions of chain length, composition, stereochemistry, and sequence. Nature has evolved a complex system for polypeptide synthesis that gives essentially complete control of chain length and monomer sequence. Using the natural protein biosynthesis machinery to produce protein polymers provides not only a unique opportunity to study the effects of such molecular characteristics on material properties, but also the possibility of readily incorporating bioactive domains into protein-based materials.

The objective of this thesis work was to expand upon the set of amino acids available for incorporation into proteins *in vivo* and to explore applications of the novel chemistries and physical properties provided by the new analogs.

Chapter 2 describes the incorporation of new unsaturated analogues of isoleucine, the alkene 2-amino-3-methyl-4-pentenoic acid and the alkyne 2-amino-3-methyl-4-pentynoic acid, by the wild type *E. coli* biosynthetic apparatus. Incorporation was found to be sensitive to side chain stereochemistry in the case of the alkene analog; the translational activity of the pairs of enantiomers (SS, RR and SR, RS) were markedly different. We concluded that, although the SS-isomer is a good analogue, the SR-isomer is not incorporated into proteins by this expression host.

Chapter 3 focuses on the incorporation of a fluorine-containing noncanonical amino acid, 5,5,5-trifluoroisoleucine, into artificial extracellular matrix proteins. The fluorinated proteins displayed altered solubility phase behavior and were more resistant to degradation by the physiologically relevant protease elastase, yet retained the ability to adhere endothelial cells in a sequence specific manner.

Chapter 4 describes the incorporation of the photoreactive noncanonical analog *p*-azidophenylalanine into artificial extracellular matrix proteins. Films of the azide-containing proteins were crosslinked upon short exposure to ultraviolet radiation. Using simple patterned masks, we demonstrated the ability to pattern protein films by only exposing certain regions. When protein patters were produced on a non-adhesive background, endothelial cells selectively adhered to the protein regions to create stable cell patterns.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

2xYT	two times yeast/tryptone medium
5TFI	5,5,5-trifluoroisoleucine
aaRS	aminoacyl-tRNA synthetase (*)
aECM	artificial extracellular matrix protein
ATP-PP$_i$	adenosine triphosphate-pyrophosphate
βME	β-mercaptoethanol
BSA	bovine serum albumin
CAI	cell adhesion index
Cbz	*N*-benzyloxycarbonyl
CD	circular dichroism
Cy2	cyanine dye 2
DAPI	4′, 6 diamindine-2-phenyl indole
DMSO	dimethylsulfoxide
dpi	dots per inch
DPN	dip-pen nanolithography
E	elastic modulus
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
E-Ile	2-amino-3-methyl-4-pentenoic acid
E. coli	*Escherichia coli*
FTIR	Fourier transform infrared spectroscopy
HLE	human leukocyte elastase
^{1}H-NMR	proton nuclear magnetic resonance (spectroscopy)
HUVEC	human umbilical vein endothelial cells
IleRS	isoleucyl-tRNA synthetase
IPTG	isopropyl-β-D-thiogalactopyranoside
L-allo-Ile	(2S, 3R) 2-amino-3-methyl-4-pentanoic acid

LCST lower critical solution temperature

M9 or M9AA minimal medium

MALDI-MS matrix-assisted laser desorption ionization-mass spectrometry

mDHFR murine dihydrofolate reductase

NEB New England Biolabs

OD optical density

PBS phosphate-buffered saline

PEG poly(ethylene glycol)

PheRS phenylalanyl-tRNA synthetase

PMSF phenylmethylsulfonyl fluoride

***p*N$_3$Phe** *para*-azidophenylalanine

PP$_i$ sodium pyrophosphate

SDS-PAGE sodium dodecyl sulfate - polyacrylamide gel electrophoresis

Y-Ile 2-amino-3-methyl-4-pentynoic acid

Common abbreviations for the twenty canonical amino acids

amino acid	3-letter	1-letter	amino acid	3-letter	1-letter
alanine	Ala	A	leucine	Leu	L
arginine	Arg	R	lysine	Lys	K
aspartic acid	Asp	D	methionine	Met	M
asparagine	Asn	N	phenylalanine	Phe	F
cysteine	Cys	C	proline	Pro	P
glutamic acid	Glu	E	serine	Ser	S
glutamine	Gln	Q	threonine	Thr	T
glycine	Gly	G	tryptophan	Trp	W
histidine	His	H	tyrosine	Tyr	Y
isoleucine	Ile	I	valine	Val	V

**The abbreviation for an aminoacyl-tRNA synthetase specific to an amino acid is formed by placing the appropriate three-letter amino acid abbreviation before the letters RS. For example, the isoleucyl-tRNA synthetase is abbreviated IleRS.*

C h a p t e r I

INTRODUCTION

Traditional techniques of polymer synthesis produce macromolecules with statistical distributions of chain length, composition, stereochemistry, and sequence. Nature has evolved a complex system for polypeptide synthesis that gives essentially complete control of chain length and monomer sequence. Using the natural protein biosynthesis machinery to produce protein polymers provides not only a unique opportunity to study the effects of such molecular characteristics on material properties, but also the possibility of readily incorporating bioactive domains into protein-based materials.

Biomaterials produced through protein biosynthesis are limited, however, to the chemical functionality contained within the canonical set of 20 amino acids (Figure I-1). The incorporation of noncanonical amino acids makes possible the introduction of new functionality into proteins, creating the potential for novel material properties or interesting post-translational modifications. This thesis describes a series of investigations that i) expand the number of amino acids available for *in vivo* incorporation into proteins and ii) use the new physical and chemical properties provided by noncanonical amino acids to produce interesting materials.

The first step toward the goal of producing polypeptide polymers containing new amino acid functionality must be an understanding of how proteins are synthesized in nature.

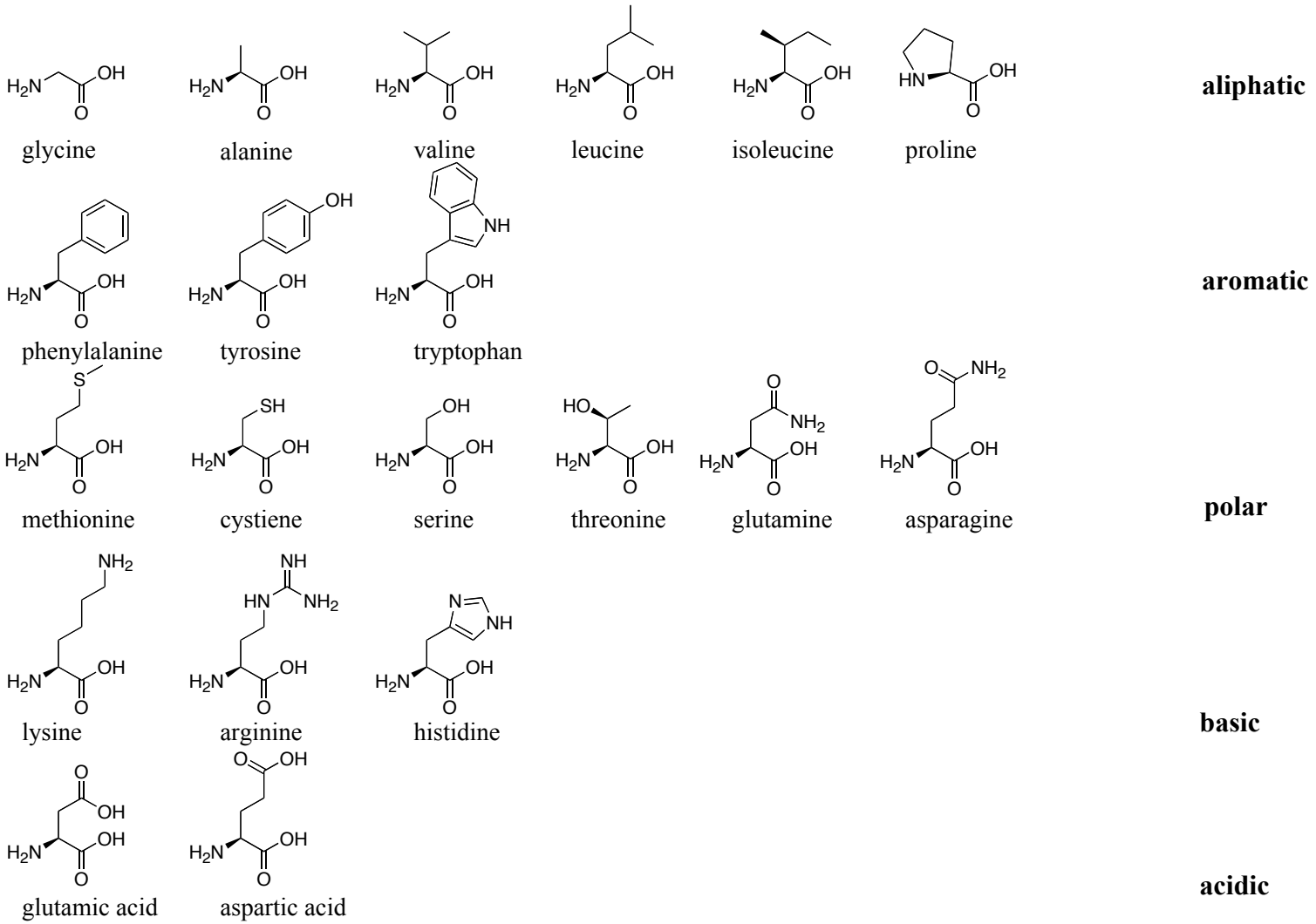


Figure I-1. The 20 canonical amino acids contain a variety of functional groups.

I.1 Protein Biosynthesis

Synthesis by the natural biosynthetic machinery involves transcription of DNA into mRNA and translation of mRNA codons into an amino acid sequence by tRNA (Figure I-2). Each tRNA contains a three-base anticodon that permits recognition of the codon(s) for the appropriate amino acid and, at the opposite end of the molecule, an adenine that serves as the attachment point for the amino acid. Amino acids are charged onto the tRNA by aminoacyl-tRNA synthetases (aaRSs); the aminoacyl-tRNAs are then transported to the ribosome by elongation factor Tu (EF-Tu). The ribosome, a multi-unit protein/RNA complex, guides the association of aminoacyl-tRNAs with their cognate mRNA codons, and a peptide bond is formed between the growing protein chain and the

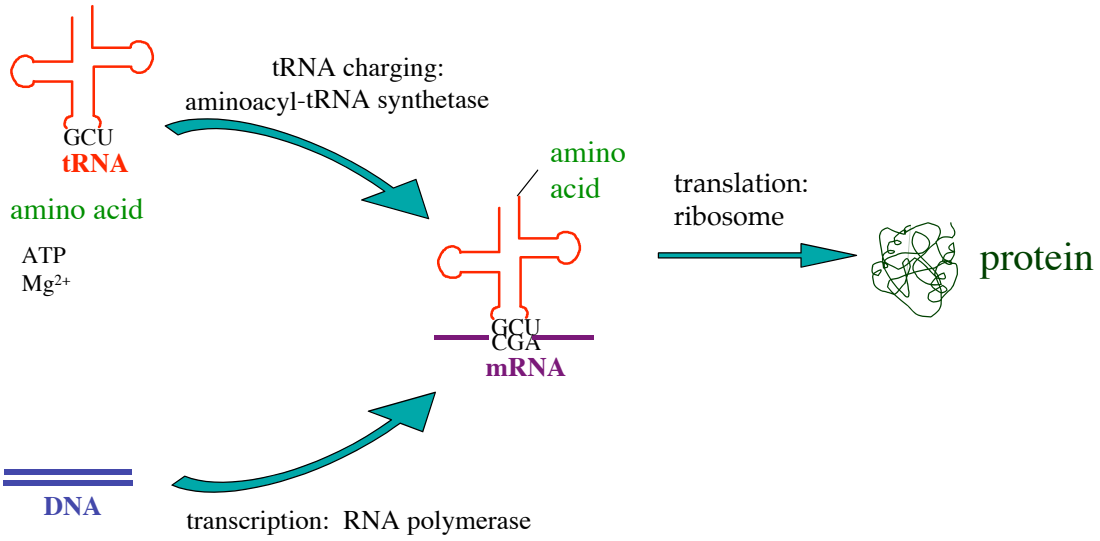


Figure I-2. During protein biosynthesis the genetic information encoded in DNA is transcribed into mRNA. The message is translated into an amino acid sequence in the ribosome where mRNA codons are "read" by the appropriate tRNA molecule that has been charged with its cognate amino acid by an aminoacyl-tRNA synthetase.

amino acid charged to the tRNA. The selectivity of this process for the appropriate amino acid is responsible for translational fidelity and is an important consideration in the attempt to incorporate noncanonical amino acids *in vivo*.

I.1.1 Fidelity of Protein Biosynthesis

There are a number of points during protein biosynthesis that provide an opportunity to assure that the correct amino acid is incorporated into the protein chain in response to a codon. The aaRSs recognize both the amino acid and one or more of its cognate tRNAs, EF-Tu binds the aminoacyl-tRNA to deliver it to the ribosome, and the ribosome itself recognizes the aminoacyl-tRNA.

There is considerable evidence that the aaRSs are largely responsible for the fidelity of protein synthesis. The ribosome discriminates between D- and L-α-amino acids [1-3] and rejects large aromatic analogues with certain geometries [4], but is otherwise relatively insensitive to amino acid structure. Indeed, in *in vitro* translation systems in which the amino acid is first attached to tRNA by chemical misacylation to bypass the aaRSs [5], the ribosome has been shown to accept and incorporate dozens of noncanonical amino acids into growing protein chains. Incorporated analogues include α-amino acids with side chains that are much larger and/or very chemically distinct from the canonical amino acids [6-8] as well as non-α-amino acids [9], including α-hydroxy acids [1, 10] and N-methyl amino acids [11, 12].

In the case of EF-Tu, it is known that the strength of binding of aminoacyl-tRNAs is optimized when the amino acid is attached to its cognate tRNA [13]. Yet, again, EF-Tu is promiscuous enough to transport to the ribosome a large number of misacylated tRNAs

produced both *in vitro* [14] and *in vivo* [15]. The majority of the burden of choosing the correct amino acid from the available intracellular pool rests with the aaRSs.

Each aaRS recognizes both an amino acid and its cognate tRNA(s). Although each aaRS has evolved to recognize a particular amino acid, common structural and functional domains have been identified (for review, see [16]). The aaRSs first catalyze the activation of the carboxyl group of the amino acid by reaction with adenosine triphosphate (ATP) to produce an aminoacyl adenylate. The activated ester of the aminoacyl adenylate reacts with the terminal hydroxyl group of tRNA to produce aminoacyl-tRNA. In some cases, the aaRS also catalyzes hydrolysis of improperly charged and/or activated amino acids.

Once the tRNA is aminoacylated by the aaRS, it is transported by EF-Tu to the ribosome where synthesis of the polypeptide chain is completed. Accurate decoding of genetic information is accomplished by the protein-mediated matching of amino acids with oligonucleotide sequence.

I.2 Protein Polymers

To the synthetic polymer chemist, an attractive feature of the protein biosynthesis machinery is its ability to precisely control the sequence of the growing polymer chain. In nature, such a fine degree of molecular control permits the production of proteins with invariable sequence and the ability to fold into well-defined, functional three-dimensional structures. Using the tools nature has developed provides the chemist with a way to design and produce protein-based polymers with controlled molecular architecture.

Early work in the Tirrell laboratory demonstrated the ability to introduce novel material properties through the precise control of polymer architecture afforded by the biosynthetic machinery. The rod-like polymer poly(γ-benzyl-L-glutamate) (PBLG) has been studied for its ability to form liquid crystalline phases; when produced with low polydispersity (~1.2) by standard synthetic techniques such as ring-opening polymerization of α-glutamic acid-N-carboxyanhydride, PBLG forms nematic liquid crystalline phases [17]. However, when PBLG is prepared through post-translational modification of bacterially produced poly(L-glutamic acid), a more ordered smectic phase is formed; mass spectroscopy of such samples show the chain population is of one uniform length [17].

Genetically templated synthesis also provides a straightforward method to include bioactive domains in the polymer structure. For example, the Tirrell laboratory has produced a family of proteins designed to mimic the natural extracellular matrix [18]. These modular constructs contain repeating blocks of protein sequence, one block derived from the structural protein elastin to impart appropriate mechanical properties and another block derived from known cell-adhesive sequences of the natural extracellular matrix protein fibronectin. Such proteins, when crosslinked by various methods, possess moduli within the range of native elastin [19,20] and adhere human umbilical vein endothelial cells (HUVEC) in a sequence-specific manner [21]. Studies involving these artificial extracellular matrix proteins (aECMs) will be the focus of Chapters 3 – 5 of this thesis.

Although *in vivo* synthesis of protein polymers has been shown to be a powerful technique, a clear limitation is the availability of only 20 monomers — the 20 canonical amino acids — with a limited range of functionality (Figure I-1). Expanding the number of monomers beyond these 20 would provide access to new chemical reactivity not available in wild type proteins.

I.3 Strategies for the Incorporation of Noncanonical Amino Acids

The protein biosynthetic machinery makes mistakes; error frequencies for misreading a codon and incorporating the incorrect canonical amino acid are ~10^{-4} [22]. It has also long been recognized that certain noncanonical amino acids, e.g., selenomethionine [23], can infiltrate the biosynthetic machinery and be incorporated into proteins in place of a structurally similar amino acid. Work in several laboratories has exploited this promiscuity to incorporate amino acids with a variety of functional groups not contained within the canonical set [24-32].

To further expand the set of functional groups available to protein engineers, several groups have been developing techniques to incorporate noncanonical amino acids, both in bacterial and mammalian cells as well as in cell-free *in vitro* translation systems. Figure I-3 schematically illustrates the major strategies being applied for the incorporation of noncanonical amino acids into proteins *in vitro* and *in vivo*.

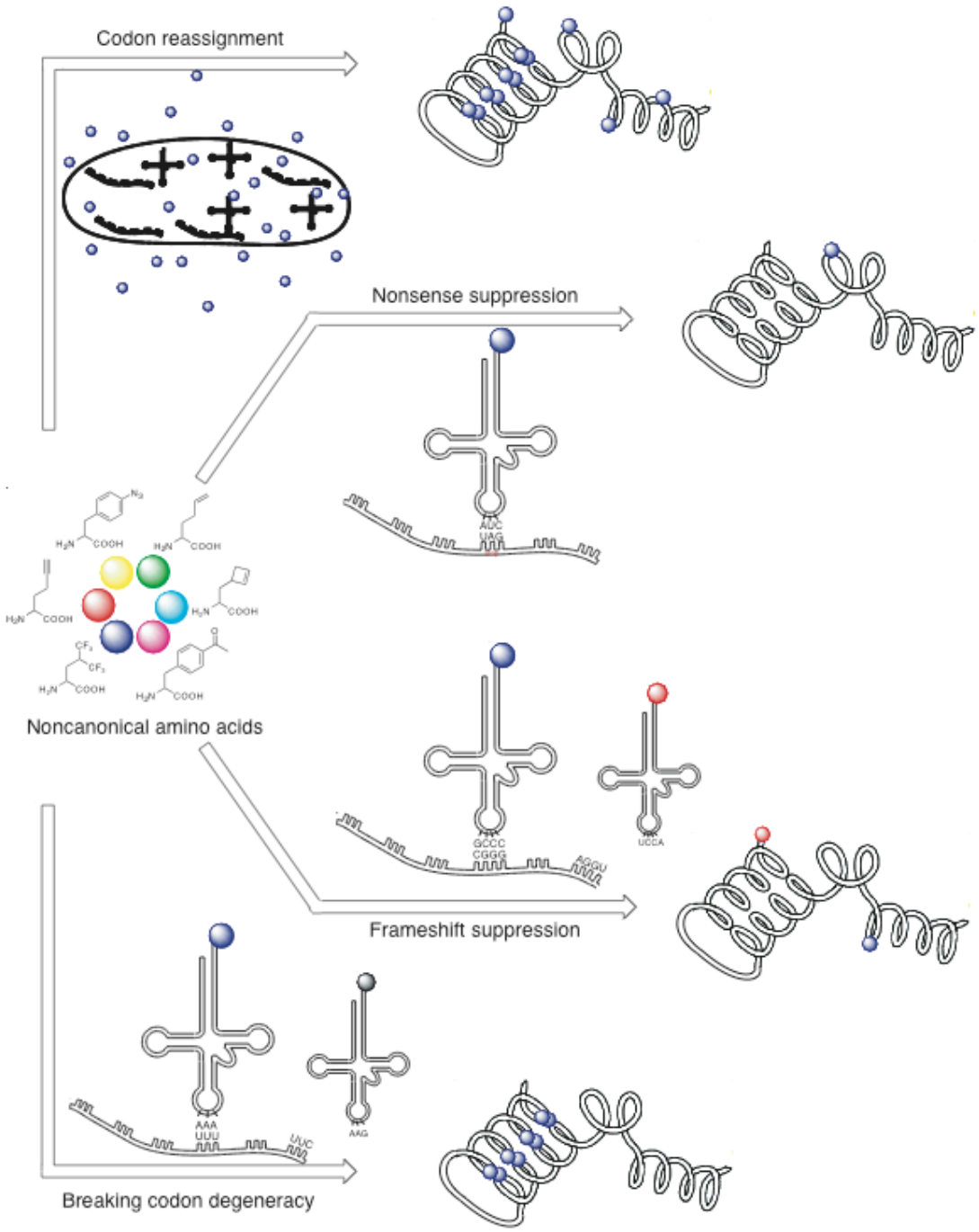


Figure I-3. Schematic representation of the various strategies for incorporating noncanonical amino acids into proteins. This figure appeared in [15].

I.3.1 In vitro *incorporation*

In the late 1980s Chamberlin [33] and Schultz [34] introduced an *in vitro* strategy for incorporation of noncanonical amino acids through nonsense suppression by chemically misacylated tRNA. Many groups have since used this technology for site-specific incorporation to study protein structure and function.

The nonsense suppression strategy uses translational read-through by a chemically misacylated suppressor tRNA to incorporate noncanonical amino acids site-specifically in response to a nonsense codon, usually the amber stop codon (TAG). Because of limitations in suppressor efficiencies, protein yields are generally low (1 – 10 μg/mL) [35], but the precise control of the placement of the noncanonical analogue makes this a powerful tool for studies of protein structure and function.

For example, Pollitt and coworkers have employed site-specific incorporation *in vitro* to cage an aspartic acid side chain of a protein ($p21^{ras}$) as its *o*-nitrobenzyl ester, allowing them to photochemically control its interaction with a protein partner (p120-GAP) [36]. Koh and coworkers have investigated the role of specific backbone amide linkages in T4 lysozyme through substitution with ester bonds by incorporation of α-hydroxy acids [10].

Site-specific incorporation has also been accomplished using codons other than stop codons. Frameshift suppression of four-base codons has been used by Sisido and coworkers to incorporate a fluorophore-quencher pair at selected positions in strepavidin

[37], and use of unnatural nucleosides *iso*-C and *iso*-G enabled the Hecht group to incorporate iodotyrosine in response to the "65th codon" [38].

I.3.2 In vivo *incorporation*

I.3.2.1 Residue-specific incorporation (codon reassignment)

By reassigning all of the codons for a particular amino acid to a close structural analogue, the noncanonical analogue can be incorporated at multiple sites throughout the protein. Using this technique, one can significantly change the physical properties of a protein [28, 39, 40] or provide for multiple reactive sites for chemical modification, such as labeling [25, 41] or crosslinking (Chapter 4).

Often codon reassignment is accomplished simply by starving an auxotrophic *Escherichia coli* (*E. coli*) strain for a canonical amino acid in medium supplemented with a close structural analogue. Tang and coworkers [40] employed this strategy to replace up to 92% of the leucine resides in the leucine-zipper protein A1 with trifluoroisoleucine. The fluorinated leucine zippers displayed a 13°C increase in melting temperature (T_m) as well as increased resistance to chemical denaturation. Kiick and coworkers [25] also used this method to incorporate the noncanonical amino acid azidohomoalanine into a target protein, murine dihydrofolate reductase (mDHFR), which they then chemoselectively labeled through Staudinger ligation with a phosphine bearing an antigenic peptide.

Access to a wider range of noncanonical amino acids for residue-specific incorporation is provided through alterations of the wild-type biosynthetic machinery. Overexpression of the appropriate synthetase can improve the incorporation of poorly activated analogues [39,42]. A more general strategy involves mutating the aaRS to alter its substrate specificity at either the synthetic [24,26,43] or hydrolytic [44,45] active site. For example, Doring and coworkers demonstrated replacement of valine by the noncanonical amino acid aminobutyrate in a strain harboring valyl-tRNA synthetase (ValRS) with multiple mutations in the editing site [45].

A previously characterized mutant *E. coli* phenylalanyl-tRNA synthetase (PheRS) [24,46], with an enlarged active site resulting from the mutation of residue 294 from alanine to glycine, accepts a wide variety of phenylalanine analogues not incorporated by wild-type *E. coli* hosts, including the photoreactive amino acid *para*-azidophenylalanine [26]. Photocrosslinking of artificial proteins through this noncanonical amino acid is described in Chapter 4 of this thesis.

I.3.2.2 Site-specific incorporation

Several research groups have developed technologies to site-specifically incorporate noncanonical amino acids *in vivo*. Lester and coworkers first demonstrated site-specific incorporation *in vivo* through injection of chemically misacylated tRNA into *Xenopus* oocytes [14] and have more recently expanded this strategy to mammalian cells [47].

Furter [48] developed a successful system for site-specific incorporation in *E. coli* by importing a yeast suppressor tRNA and yeast aminoacyl-tRNA synthetase to generate

a "21st pair." Because the native *E. coli* $tRNA^{Phe}$/PheRS pair was incapable of activating *p*-fluorophenylalanine, this noncanonical analogue was selectively incorporated by the heterologous 21st pair in response to amber stop codons.

Schultz and coworkers have expanded and improved upon this strategy. Through directed evolution of aaRSs that show high specificity toward a noncanoncial amino acid [49], they have created *E. coli* hosts with 21st pairs incorporating *p*-acetylphenylalanine [50] and benzophenone [51], among others. They have also generated an organism with a novel 21st amino acid by inserting an aaRS evolved to incorporate *p*-aminophenylalanine and the biosynthetic genes for the amino acid into an *E. coli* host [52]. Use of orthogonal 21st pairs has recently been expanded to *Saccharomyces. cerevisiae* [53] and Chinese hamster ovary cells [54].

I.3.2.3 Multiple site-specific incorporation

Kwon and coworkers have described a system that takes advantage of the degeneracy of the genetic code to "reassign" the phenylalanine wobble codon UUU to the noncanonical amino acid 2-napthylalanine [55]. This technique is complementary to other missense strategies such as nonsense and frameshift suppression, but offers the unique ability to efficiently incorporate a noncanonical amino acid at multiple pre-selected sites, e.g., throughout a single protein domain.

Using the strategies described in Section I.3.2, to date the Tirrell laboratory has successfully incorporated the set of amino acids shown in Figure I-4 into proteins *in vivo*.

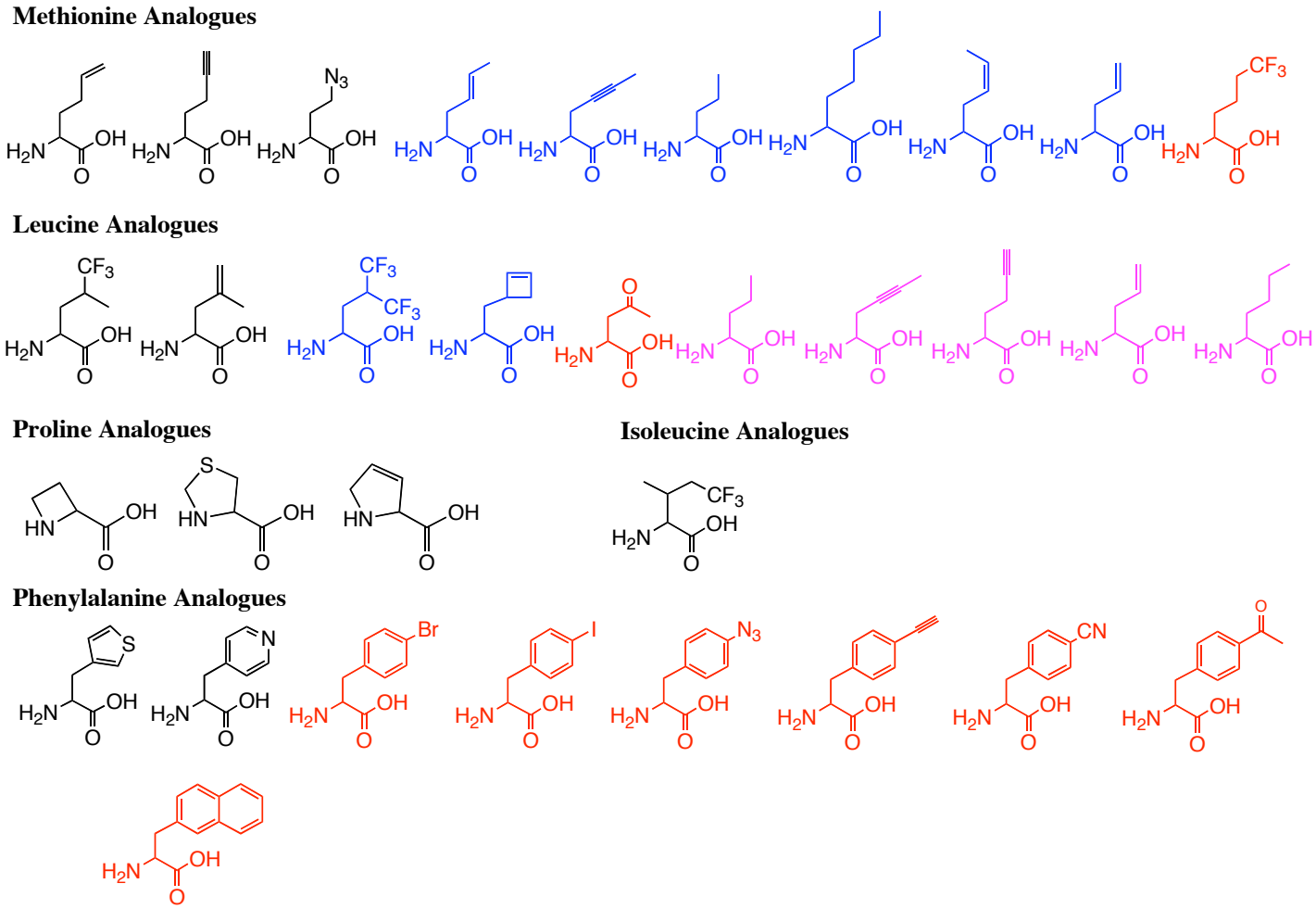


Figure I-4. Noncanonical amino acids incorporated into proteins *in vivo* by the Tirrell laboratory. Amino acids in black are accepted by the wild type *E. coli* biosynthetic machinery; those in blue require overexpression of the wild type aaRS; those in red require an active site mutant aaRS; and those in purple require an editing site mutant aaRS.

I.4 Thesis Objective

The objective of this thesis work was to expand upon the set of amino acids available for incorporation into proteins *in vivo* and to explore applications of the novel chemistries and physical properties provided by the new analogues.

Chapter 2 describes the incorporation of new, unsaturated analogues of isoleucine, the alkene 2-amino-3-methyl-4-pentenoic acid (E-Ile) and the alkyne 2-amino-3-methyl-4-pentynoic acid (Y-Ile), by the wild type *E. coli* biosynthetic apparatus. The IleRS was found to be sensitive to sidechain stereochemistry in the case of the alkene analogue; the translational activity of the pairs of enantiomers (SS, RR-E-Ile and SR, RS-E-Ile) are markedly different. We conclude that although SS-E-Ile is a good substrate for the IleRS, SR-E-Ile is not incorporated into proteins by wild type *E. coli* translational machinery. At least one stereoisomer of the alkyne analogue of Ile is also translationally active, although the level of stereochemical purity of the amino acid samples was not sufficient to determine whether a similar stereochemical discrimination existed for Y-Ile.

Chapter 3 focuses on the incorporation of a fluorinated noncanonical amino acid, 5,5,5-trifluoroisoleucine (5TFI), into artificial extracellular matrix proteins. The fluorinated aECMs displayed altered phase behavior and were more resistant to degradation by the physiologically relevant protease elastase, yet retained the ability to adhere endothelial cells in a sequence specific manner.

Chapter 4 describes the incorporation of the photoreactive noncanonical analogue *p*-azidophenylalanine (*p*N_3Phe) into artificial extracellular matrix proteins using an *E. coli* host expressing a mutant PheRS (A294G) in which the active site has been expanded

to accommodate larger amino acids. Films of the azide-containing aECMs were crosslinked upon short (~30 s) exposure to ultraviolet radiation at 365 nm. Using simple patterned masks, we demonstrated the ability to pattern protein films by only exposing certain regions. When protein patterns are produced on a non-adhesive background, endothelial cells selectively adhere to the protein regions to create stable cell patterns.

Chapter 5 describes progress toward the cloning of a new aECM construct containing a very robust cell-adhesive domain as well as regular phenylalanine sites for incorporation of *p*N_3Phe.

I.5 References

1. Bain JD, Diala ES, Glabe CG, Wacker DA, Lyttle MH, Dix TA, Chamberlin AR: **Site-Specific Incorporation of Nonnatural Residues During Invitro Protein-Biosynthesis with Semisynthetic Aminoacyl-Transfer Rnas**. *Biochemistry* 1991, **30**:5411-5421.

2. Roesser JR, Xu C, Payne RC, Surratt CK, Hecht SM: **Preparation of Misacylated Aminoacyl-Transfer Rna Phes Useful as Probes of the Ribosomal Acceptor Site**. *Biochemistry* 1989, **28**:5185-5195.

3. Heckler TG, Roesser JR, Xu C, Chang PI, Hecht SM: **Ribosomal-Binding and Dipeptide Formation by Misacylated Transfer Rnaphes**. *Biochemistry* 1988, **27**:7254-7262.

4. Hohsaka T, Sato K, Sisido M, Takai K, Yokoyama S: **Adaptability of Nonnatural Aromatic-Amino-Acids to the Active-Center of the Escherichia-Coli Ribosomal a-Site**. *Febs Letters* 1993, **335**:47-50.

5. Heckler TG, Chang LH, Zama Y, Naka T, Chorghade MS, Hecht SM: **T4 Rna Ligase Mediated Preparation of Novel Chemically Misacylated Transfer-Rna Phes**. *Biochemistry* 1984, **23**:1468-1473.

6. Sisido M, Hohsaka T: **Extension of protein functions by the incorporation of nonnatural amino acids**. *Bulletin of the Chemical Society of Japan* 1999, **72**:1409-1425.

7. Gilmore MA, Steward LE, Chamberlin AR: **Incorporation of noncoded amino acids by in vitro protein biosynthesis**. In *Implementation and Redesign of Catalytic Function in Biopolymers*; 1999:77-99. Topics in Current Chemistry, vol 202.

8. Mendel D, Cornish VW, Schultz PG: **Site-Directed Mutagenesis with an Expanded Genetic-Code**. *Annual Review of Biophysics and Biomolecular Structure* 1995, **24**:435-462.

9. Eisenhauer BM, Hecht SM: **Site-specific incorporation of (aminooxy)acetic acid into proteins**. *Biochemistry* 2002, **41**:11472-11478.

10. Koh JT, Cornish VW, Schultz PG: **An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis**. *Biochemistry* 1997, **36**:11314-11322.

11. Ellman JA, Mendel D, Schultz PG: **Site-Specific Incorporation of Novel Backbone Structures into Proteins**. *Science* 1992, **255**:197-200.

12. Bain JD, Wacker DA, Kuo EE, Chamberlin AR: **Site-Specific Incorporation of Nonnatural Residues into Peptides - Effect of Residue Structure on Suppression and Translation Efficiencies**. *Tetrahedron* 1991, **47**:2389-2400.

13. LaRiviere FJ, Wolfson AD, Uhlenbeck OC: **Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation**. *Science* 2001, **294**:165-168.

14. Nowak MW, Kearney PC, Sampson JR, Saks ME, Labarca CG, Silverman SK, Zhong W, Thorson J, Abelson JN, Davidson N, Schultz PG, Dougherty DA, Lester HA: **Nicotinic Receptor-Binding Site Probed with Unnatural Amino-Acid-Incorporation in Intact-Cells**. *Science* 1995, **268**:439-442.

15. Link AJ, Mock ML, Tirrell DA: **Non-canonical amino acids in protein engineering**. *Current Opinion in Biotechnology* 2003, **14**:603-609.

16. Woese CR, Olsen GJ, Ibba M, Soll D: **Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process**. *Microbiology and Molecular Biology Reviews* 2000, **64**:202-236.

17. Yu SJM, Soto CM, Tirrell DA: **Nanometer-scale smectic ordering of genetically engineered rodlike polymers: Synthesis and characterization of monodisperse derivatives of poly(gamma-benzyl alpha,L-glutamate)**. *Journal of the American Chemical Society* 2000, **122**:6552-6559.

18. Panitch A, Yamaoka T, Fournier MJ, Mason TL, Tirrell DA: **Design and biosynthesis of elastin-like artificial extracellular matrix proteins containing periodically spaced fibronectin CS5 domains**. *Macromolecules* 1999, **32**:1701-1703.

19. Di Zio K, Tirrell DA: **Mechanical properties of artificial protein matrices engineered for control of cell and tissue behavior**. *Macromolecules* 2003, **36**:1553-1558.

20. Welsh ER, Tirrell DA: **Engineering the extracellular matrix: A novel approach to polymeric biomaterials. I. Control of the physical properties of artificial protein matrices designed to support adhesion of vascular endothelial cells**. *Biomacromolecules* 2000, **1**:23-30.

21. Heilshorn SC, DiZio KA, Welsh ER, Tirrell DA: **Endothelial cell adhesion to the fibronectin CS5 domain in artificial extracellular matrix proteins**. *Biomaterials* 2003, **24**:4245-4252.

22. Kurland CG: **Translational Accuracy and the Fitness of Bacteria**. *Annual Review of Genetics* 1992, **26**:29-50.

23. Cohen GN, Cowie DB: **Remplacement Total De La Methionine Par La Selenomethionine Dans Les Proteines Descherichia-Coli**. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* 1957, **244**:680-683.

24. Ibba M, Hennecke H: **Relaxing the substrate-specificity of an aminoacyl-transfer-rna synthetase allows in-vitro and in-vivo synthesis of proteins containing unnatural amino-acids**. *Febs Letters* 1995, **364**:272-275.

25. Kiick KL, Saxon E, Tirrell DA, Bertozzi CR: **Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**:19-24.

26. Kirshenbaum K, Carrico IS, Tirrell DA: **Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues**. *Chembiochem* 2002, **3**:235-237.

27. Tang Y, Ghirlanda G, Vaidehi N, Kua J, Mainz DT, Goddard WA, DeGrado WF, Tirrell DA: **Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine**. *Biochemistry* 2001, **40**:2790-2796.

28. Wang P, Tang Y, Tirrell DA: **Incorporation of trifluoroisoleucine into proteins in vivo**. *Journal of the American Chemical Society* 2003, **125**:6900-6906.

29. Bilgicer B, Fichera A, Kumar K: **A coiled coil with a fluorous core**. *Journal of the American Chemical Society* 2001, **123**:4393-4399.

30. Bae JH, Rubini M, Jung G, Wiegand G, Seifert MHJ, Azim MK, Kim JS, Zumbusch A, Holak TA, Moroder L, Huber R, Budisa N: **Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins**. *Journal of Molecular Biology* 2003, **328**:1071-1081.

31. Bann JG, Pinkner J, Hultgren SJ, Frieden C: **Real-time and equilibrium F-19-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**:709-714.

32. Bae JH, Alefelder S, Kaiser JT, Friedrich R, Moroder L, Huber R, Budisa N: **Incorporation of beta-selenolo 3,2-b pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography**. *Journal of Molecular Biology* 2001, **309**:925-936.

33. Bain JD, Glabe CG, Dix TA, Chamberlin AR, Diala ES: **Biosynthetic Site-Specific Incorporation of a Non-Natural Amino-Acid into a Polypeptide**. *Journal of the American Chemical Society* 1989, **111**:8013-8014.

34. Noren CJ, Anthonycahill SJ, Griffith MC, Schultz PG: **A General-Method for Site-Specific Incorporation of Unnatural Amino-Acids into Proteins**. *Science* 1989, **244**:182-188.

35. Hendrickson TL, de Crecy-Lagard V, Schimmel P: **Incorporation of nonnatural amino acids into proteins**. *Annual Review of Biochemistry* 2004, **73**:147-176.

36. Pollitt SK, Schultz PG: **A photochemical switch for controlling protein-protein interactions**. *Angewandte Chemie-International Edition* 1998, **37**:2104-2107.

37. Taki M, Hohsaka T, Murakami H, Taira K, Sisido M: **Position-specific incorporation of a fluorophore-quencher pair into a single streptavidin through orthogonal four-base codon/anticodon pairs**. *Journal of the American Chemical Society* 2002, **124**:14586-14590.

38. Bain JD, Switzer C, Chamberlin AR, Benner SA: **Ribosome-Mediated Incorporation of a Nonstandard Amino-Acid into a Peptide through Expansion of the Genetic-Code**. *Nature* 1992, **356**:537-539.

39. Tang Y, Tirrell DA: **Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host**. *Journal of the American Chemical Society* 2001, **123**:11089-11090.

40. Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF, Tirrell DA: **Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability**. *Angewandte Chemie-International Edition* 2001, **40**:1494-1496.

41. Link AJ, Tirrell DA: **Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition**. *Journal of the American Chemical Society* 2003, **125**:11164-11165.

42. Kiick KL, van Hest JCM, Tirrell DA: **Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host**. *Angewandte Chemie-International Edition* 2000, **39**:2148-2152.

43. Sharma N, Furter R, Kast P, Tirrell DA: **Efficient introduction of aryl bromide functionality into proteins in vivo**. *Febs Letters* 2000, **467**:37-40.

44. Tang Y, Tirrell DA: **Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo**. *Biochemistry* 2002, **41**:10635-10645.

45. Doring V, Mootz HD, Nangle LA, Hendrickson TL, de Crecy-Lagard V, Schimmel P, Marliere P: **Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway**. *Science* 2001, **292**:501-504.

46. Ibba M, Kast P, Hennecke H: **Substrate-specificity is determined by amino-acid binding pocket size in Escherichia-Coli phenylalanyl-transfer-rna synthetase**. *Biochemistry* 1994, **33**:7107-7112.

47. Monahan SL, Lester HA, Dougherty DA: **Site-specific incorporation of unnatural amino acids into receptors expressed in mammalian cells**. *Chemistry & Biology* 2003, **10**:573-580.

48. Furter R: **Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in Escherichia coli**. *Protein Science* 1998, **7**:419-426.

49. Wang L, Brock A, Herberich B, Schultz PG: **Expanding the genetic code of Escherichia coli**. *Science* 2001, **292**:498-500.

50. Wang L, Zhang ZW, Brock A, Schultz PG: **Addition of the keto functional group to the genetic code of Escherichia coli**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**:56-61.

51. Chin JW, Martin AB, King DS, Wang L, Schultz PG: **Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**:11020-11024.

52. Mehl RA, Anderson JC, Santoro SW, Wang L, Martin AB, King DS, Horn DM, Schultz PG: **Generation of a bacterium with a 21 amino acid genetic code**. *Journal of the American Chemical Society* 2003, **125**:935-939.

53. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang ZW, Schultz PG: **An expanded eukaryotic genetic code**. *Science* 2003, **301**:964-967.

54. Sakamoto K, Hayashi A, Sakamoto A, Kiga D, Nakayama H, Soma A, Kobayashi T, Kitabatake M, Takio K, Saito K, Shirouzu M, Hirao I, Yokoyama S: **Site-specific**

incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Research* 2002, **30**:4692-4699.

55. Kwon I, Kirshenbaum K, Tirrell DA: **Breaking the degeneracy of the genetic code**. *Journal of the American Chemical Society* 2003, **125**:7512-7513.

C h a p t e r I I

STEREOSELECTIVE INCORPORATION OF UNSATURATED ISOLEUCINE ANALOGUES INTO PROTEINS *IN VIVO**

II.1 Abstract

The unsaturated amino acids 2-amino-3-methyl-4-pentenoic acid (E-Ile) and 2-amino-3-methyl-4-pentynoic acid (Y-Ile) were prepared, and E-Ile was successfully separated into its SS, RR and SR, RS diastereomeric pairs. The translational activities of the SS-E-Ile, SR-E-Ile, and Y-Ile analogues were assessed using an *Escherichia coli* (*E. coli*) strain auxotrophic for isoleucine (Ile**)**. SS-E-Ile was incorporated into the test protein murine dihydrofolate reductase (mDHFR) in place of isoleucine at a rate of substitution of up to 72%, while SR-E-Ile showed no conclusive evidence of translational activity. At least one stereoisomer of Y-Ile also supported protein production, but the stereochemical purity of the amino acid samples was not sufficient to investigate stereochemical discrimination. *In vitro* ATP-PP_i exchange assays indicate that SS-E-Ile is activated by the isoleucyl-tRNA synthetase (IleRS) at a rate comparable to isoleucine; SR-E-Ile is activated approximately 100 times more slowly.

**Sections of this chapter are excerpted from a manuscript accepted to ChemBioChem by Marissa L. Mock, Thierry Michon, Jan C. M. van Hest, and David Tirrell, 2005.*

II.2 Introduction

Genetic engineering provides a tool with which one can prepare complex macromolecules possessing both precisely controlled architectures and specific catalytic or biological activity. Recent work has shown the advantages of using the biosynthetic machinery to produce new materials (for a review see reference [1]). The use of monomers other than the twenty canonical amino acids enables the introduction of new functionality into proteins, creating the potential for novel physical and chemical properties. Analogues of many of the canonical amino acids have been incorporated into proteins in *E. coli* using the wild-type biosynthetic machinery, e.g. [2, 3], while modifications of that machinery have permitted the incorporation of a still broader set of non-canonical amino acids [4-14]. Increasing the number of amino acid monomers that can be incorporated into proteins, and thereby the range of physical properties and chemistries available, requires detailed understanding of the biosynthetic apparatus.

Protein synthesis involves transcription of the information contained in DNA into mRNA and translation of the mRNA into polypeptide chains. The aminoacyl-tRNA synthetases (aaRSs) are essential to the fidelity of this process. Each aaRS selectively catalyzes the activation of the carboxylate group of the appropriate amino acid by reaction with adenosine triphosphate (ATP) to produce the aminoacyl adenylate, which reacts with terminal hydroxyl group of a cognate tRNA to produce aminoacyl-tRNA. The selectivity of the aaRSs is an important consideration in any attempt to incorporate nonnatural amino acids into proteins *in vivo*. Modifications of the aaRSs, through enlarging the active site [14-16] or decreasing editing activity [13, 17], have been shown to permit incorporation of analogues that are not usually incorporated into proteins. The

rational modification of aaRSs to allow use of a wider range of nonnatural amino acids requires an understanding of the mechanism(s) of selectivity of each individual aaRS.

The isoleucyl-tRNA synthetase (IleRS) has been well studied, in part because it must perform a significant feat of selective recognition as it discriminates its cognate amino acid isoleucine (Ile) from the natural amino acid valine (Val), which differs in chemical structure by only one methylene group. Pauling calculated that the additional binding energy contributed by the extra methylene group should at most result in a discrimination of 1/20 [18], while the erroneous substitution of Val for Ile actually occurs at a rate of about 1 in 3000 [19]. In fact, IleRS does misactivate Val (approximately 140 times more slowly than Ile [20, 21]) and later hydrolyzes the misactivated amino acid in an editing site located ~34 Å from the synthetic site of the enzyme [22-24]. Isoleucine contains two chiral centers, one at the alpha carbon and another at the beta carbon. The stereoisomer of 2-amino-3-methyl-heptanoic acid incorporated into proteins is (2S, 3S), designated Ile. L-allo-Ile (2S, 3R) has the correct configuration at the α-carbon, but the opposite configuration at the β-position. It is not incorporated into proteins, although there is evidence that it is bound and activated by IleRS [19, 25-27].

To expand further the chemistries available for the modification of proteins, we prepared the unsaturated Ile analogues (2S, 3S and 2R, 3R)-2-amino-3-methyl-4-pentenoic acid (SS, RR-E-Ile), (2S, 3R and 2R, 3S)-2-amino-3-methyl-4-pentenoic acid (SR, RS-E-Ile), and 2-amino-3-methyl-4-pentynoic acid (as a mixture of the SS, RR, SR, and RS stereoisomers) (Figure II-1), which have been shown previously to inhibit growth of *E. coli* [28]. We are especially interested in unsaturated amino acid analogues because of

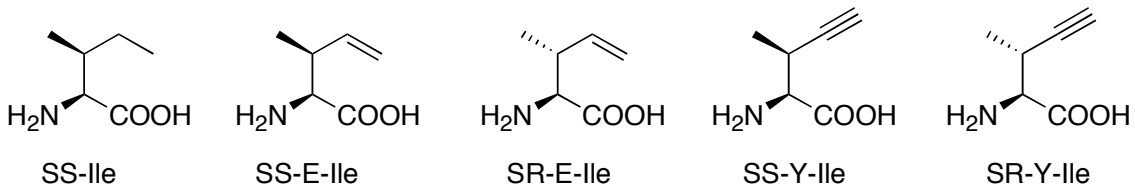


Figure II-1. Unsaturated isoleucine analogues SS-E-Ile and SR-E-Ile differ in the stereochemistry at the β-carbon. The stereoisomer of isoleucine incorporated into proteins (Ile) is 2S, 3S.

the versatile chemistry of alkenes and alkynes. For example, supramolecular structures made up of weakly hydrogen-bonded cyclic peptides can be stabilized through inter-peptide crosslinking [29] utilizing ruthenium-catalyzed ring-closing metathesis of pendant alkene moieties [30, 31] and Cu(I)-catalyzed azide-alkyne cycloaddition has been used to modify *E. coli* cells [32]. The stereoisomer pairs were evaluated with respect to incorporation into a test protein, murine dihydrofolate reductase (mDHFR), in an *E. coli* strain rendered auxotrophic for Ile. The kinetics of activation of SS and SR-E-Ile by the IleRS were also determined *in vitro* through ATP/PP_i exchange.

II.3 Methods

II.3.1 Synthesis of Analogues

II.3.1.1 General procedures

Glassware was dried at 150 °C and cooled under argon prior to use. Tetrahydrofuran (THF) was freshly distilled from sodium/benzophenone. Other reagents were used as purchased. *Cis*- and *trans*-crotyl alcohol were purchased from Chemsampco, Trenton, NJ. ^{1}H-NMR spectra were recorded on a Varian Mercury 300 MHz spectrophotometer. Column chromatography was performed on silica gel (300

Mesh, Baker) or alumina (80-200 Mesh, EM Science). Silica ($60F_{254}$ EM Science) was used for thin layer chromatography.

II.3.1.2 SS- and SR-2-amino-3-methyl-4-pentenoic acid (E-Ile)

SS- and SR-E-Ile were prepared according to Figure II-2; the stereochemistry at the β-carbon of the final product was controlled by choosing either *cis*- or *trans*-crotyl alcohol as the starting material (Figure II-3). *N*-benzyloxycarbonyl (Cbz)-protected glycine crotyl esters were prepared according to Hassner and Alexian [33], with slight modification. Cbz-glycine (17.44, 83.4 mmol) was dissolved in 100 mL THF under N_2. Either *cis*- or *trans*-crotyl alcohol (6.00 g, 83.2 mmol) was added, followed by 17.15 g (83 mmol) dicyclohexyl carbodiimide and 50 mg (6.1 mmol) dimethylaminopyridine. An exotherm was observed, and a precipitate immediately formed. The mixture was allowed to stir at room temperature for 2 days before being filtered through Celite. The

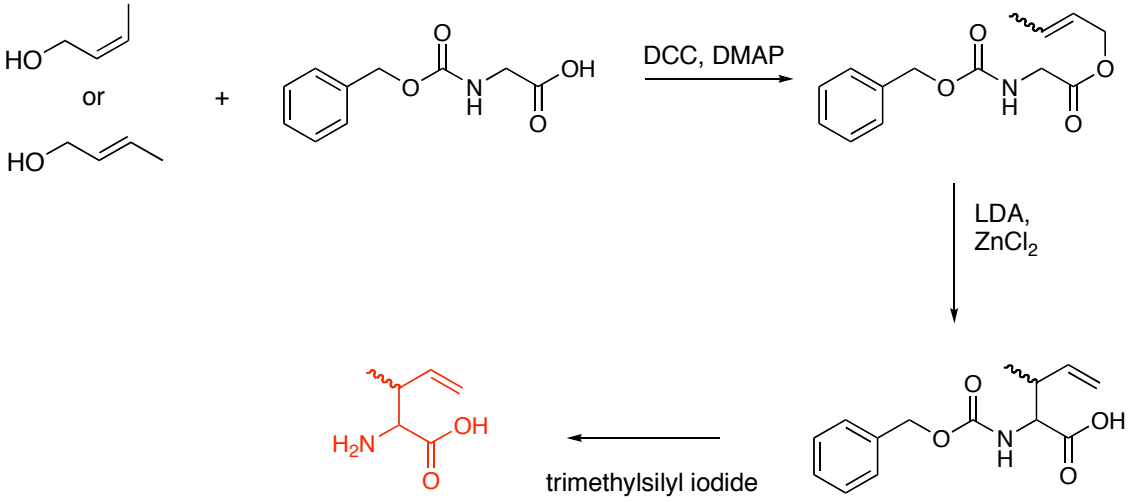


Figure II-2. Synthesis scheme for SS- and SR-E-Ile. The amine-protected crotyl esters undergo Claisen rearrangement; subsequent deprotection gives stereochemically enriched product.

solvent was evaporated, and the crude product was purified by flash chromatography (CH_2Cl_2). A viscous colorless oil was obtained by evaporation of the eluent (yield *trans*-isomer: 65 %, *cis*-isomer: 32 %). ^{1}H-NMR ($CDCl_3$): δ 1.73 (d, 3H, J = 6.9, C<u>H</u>$_3$), 3.98 (d, 2H, J = 6.4, C<u>H</u>$_2$-CO), 4.57 (d, 2H, J = 6.7, O-C<u>H</u>$_2$), 5.17 (s, 2H, Ph-C<u>H</u>$_2$), 5.56 (m, 1H, J_E = 17.4, J_Z = *10.2,* C=C<u>H</u>-CH$_3$), 5.81 (m, 1H, J_E = 17.4, J_Z = *10.2,* O-CH$_2$-C<u>H</u>=C), 7.35 (s, 5H, Ph).

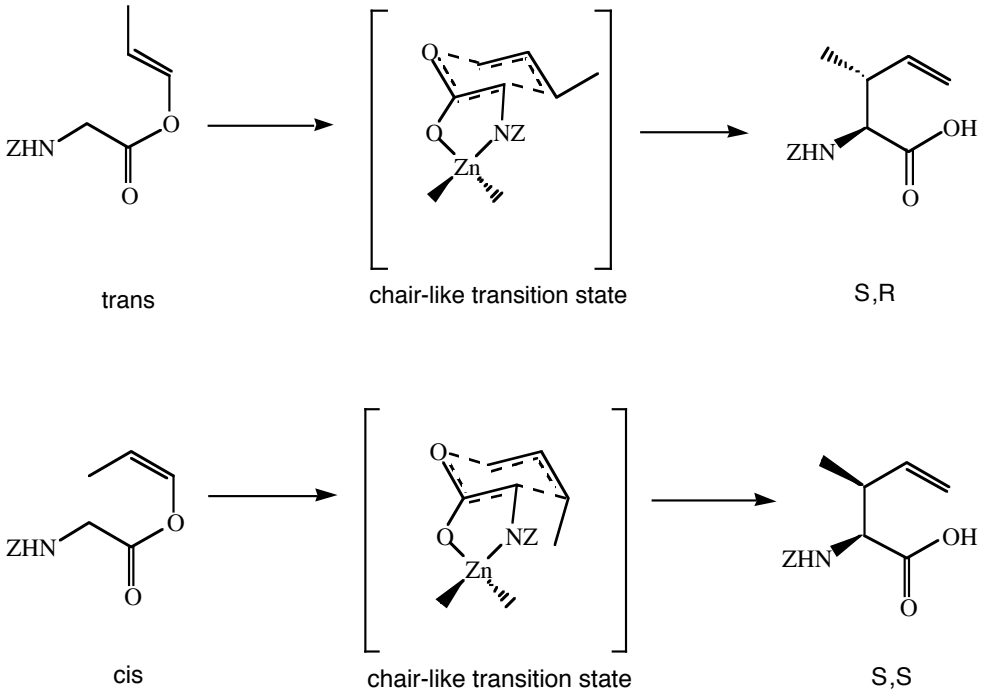


Figure II-3. Stereoselectivity in the preparation of E-Ile is a result of a preference for the chair-like transition state during the Claisen rearrangement, which determines the stereochemistry at the β-carbon in the product. The *trans*-crotyl alcohol yields SR (and RS)-E-Ile, while the *cis*-crotyl alcohol gives SS (and RR)-E-Ile.

N-benzyloxycarbonyl (Cbz)-protected glycine crotyl esters were rearranged following Kazmaier [34] to *N*-benzyloxycarbonyl-2-amino-3-methyl-4-pentenoic acid. Diisopropyl amine (6.9 mL, 45.8 mmol) was dissolved in 40 mL dry THF under argon.

The stirred solution was cooled to –20 °C and 26 mL (41.6 mmol) 1.6 M *n*-butyl lithium in *n*-hexane was added. After 20 minutes, the solution was cooled to –78 °C, and 5.4 g (20.5 mmol) Cbz-protected glycine crotyl ester (either *cis* or *trans*) in 20 mL THF and 25 mL of a 0.5 M $ZnCl_2$ solution in THF were added simultaneously over a 30 min period. The solution remained homogeneous. After an additional 30 min at –78 °C the solution was allowed to warm to room temperature. The rearrangement was monitored by thin layer chromatography (3: 7 ethyl acetate:dichloromethane). After 5 hours, 30 % (*trans*) to 50 % (*cis*) of the starting ester was still present in the reaction medium. Additional incubation time did not improve the yield of the rearrangement. The reaction was terminated by addition of 10 mL 1 M HCl. The ether phase was extracted with two 75 mL volumes of 1 M NaOH. The aqueous phase was neutralized with concentrated HCl to precipitate the acid product, which was extracted with diethyl ether (150 mL). The ether layer was dried over $MgSO_4$, and the solvent was evaporated to give a yellow oil. The acids were purified by flash chromatography (99:1 ethyl acetate:acetic acid).

Either SS, RR or RS, SR Cbz-protected 2-amino-3-methyl-4-pentenoic acid (0.8 g, 3 mmol) was dissolved in 10 mL dry $CHCl_3$ under nitrogen atmosphere. Trimethylsilyl iodide (1.2 mL, 8.8 mmol) was added. After 20 min stirring at room temperature, the reaction was quenched by addition of 1 mL methanol. The solvent was evaporated, the crude product was dissolved in 10 mL 30 % v/v acetic acid, and the solution was washed twice with 15 mL diethyl ether. The aqueous layer was evaporated, yielding a yellow oil. The SR, RS mixture of isomers crystallized upon cooling. Recrystallization from 4:1 isopropanol:water gave 0.18 g (yield 45 %) of pure amino acid (d.e.=94%). ^{1}H-NMR

(D_2O) δ: 1.08 (d, 3H, J=7.0, CH-C<u>H</u>$_3$), 2.8-2.9 (m, 1H, J_1=4.1, J_2=7.0, C<u>H</u>-CH_3,), 3.74 (d, 1H, J=4.1, NH_2-C<u>H</u>-COOH), 5.20-5.27 (m, 2H, J_1=5.9, J_2=11.2, CH=C<u>H</u>$_2$), 5.76-5.88 (m, 1H, J_1=6.2, J_2=11.6, C<u>H</u>=CH_2).

The crude SS, RR-isomer did not crystallize. It was dissolved in 6 N HCl; the hydrochloride was isolated by evaporation and dissolved in methanol. The amino acid was precipitated by slow addition of propylene oxide. The precipitate was crystallized from 4:1 isopropanol:water, yielding 0.116 g (29%) of pure SS, RR-E-Ile (d.e.=78%). ^{1}H-NMR (D_2O) δ: 1.12 (d, 3H, J=7.0, CH-C<u>H</u>$_3$), 2.75-2.85 (m, 1H, J_1=6.9, J_2=5.8, C<u>H</u>-CH_3), 3.57 (d, 1H, J=5.7, NH_2-C<u>H</u>-COOH), 5.20-5.27 (m, 2H, J_1=5.1, J_2=11.1, CH=C<u>H</u>$_2$), 5.76-5.88 (m, 1H, J_1=7.3, J_2=10.1, C<u>H</u>=CH_2).

The reaction products were recrystallized 5 times from 5:1 isopropanol:water to give SS, RR-E-Ile (d.e. = 95%) and SR, RS-E-Ile (d.e. = 98%).

II.3.1.3 <u>2-amino-3-methyl-4-pentynoic acid (Y-Ile)</u>

The synthesis scheme for Y-Ile appears in Figure II-4. Diphenylmethylene glycine ethyl ester was prepared according to O'Donnell and Polt [35]. The ^{1}H NMR spectrum was in accord with Aidene and coworkers [36].

To prepare 3-bromobutyne, 6.82 g (62 mol) PBr_3 containing 0.1 mg hydroquinone and a solution of 10 g (0.14 mol) 3-butyn-2-ol in 1 mL dry pyridine were added dropwise simultaneously over a 4-hour period to a 50 mL round bottom flask cooled to -15 °C, maintained under argon, and equipped with a stirrer, an addition funnel, and a gas inlet. After an additional 30 min of reaction, 20 mL of cold water were cautiously added to stop the reaction. The mixture was extracted with ether (3 x 20 mL). The combined ether layers were washed consecutively with water (3 x 20 mL), saturated sodium bicarbonate

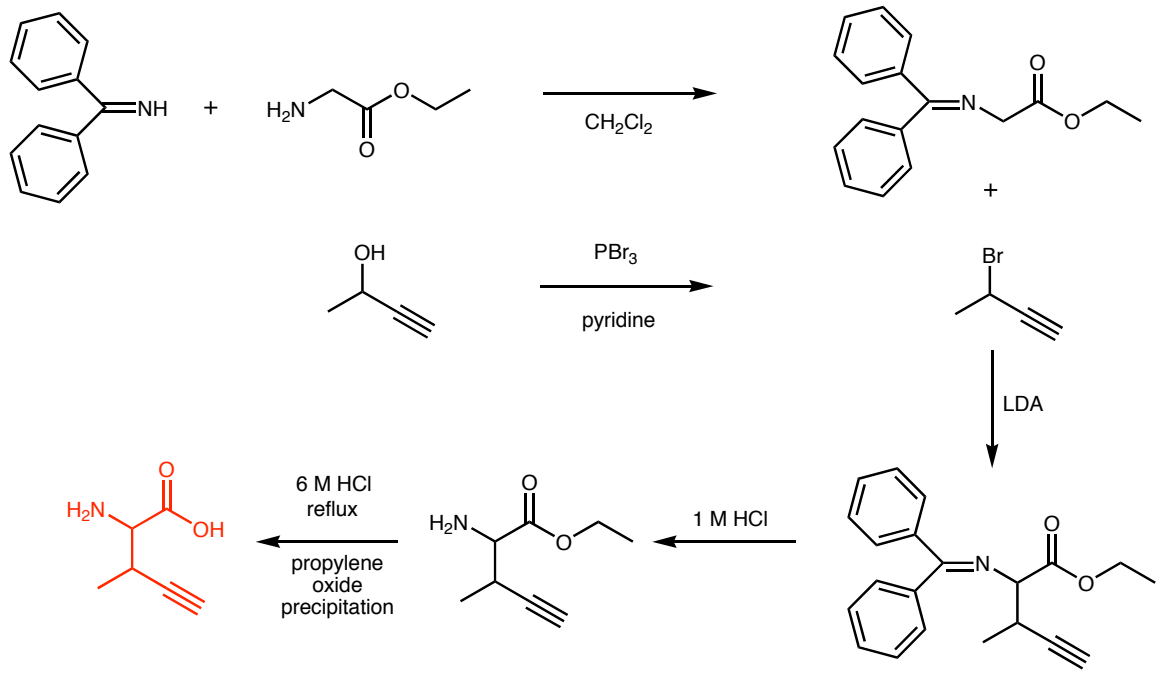


Figure II-4. Y-Ile was prepared from diphenylmethylene glycine ethyl ester, which was alkylated with 3-bromobutyne. The alkylated product was deprotected to yield a mixture of stereoisomers of Y-Ile.

(3 x 25 mL), and saturated sodium chloride (2 x 50 mL). The ether extract was dried with $MgSO_4$ and filtered. The dry ether phase was immediately distilled, affording 8.1 g (45%) of 3-bromobutyne. ^{1}H-NMR ($CDCl_3$): δ 1.90 (d, 3H, J = 7.1, C<u>H</u>$_3$-CHBr), 2.63 (d, 1H, CHBr-CC<u>H</u>), 4.55 (m, 1H, J = 7.1, CH_3-C<u>H</u>Br) ppm.

A solution of 1.16 mL (8.22 mmol) diisopropylamine in 15 mL dry tetrahydrofuran (THF) was cooled to –20 °C under an argon atmosphere. A 1.6 M solution of *n*-butyl lithium in *n*-hexane (5.13 mL, 8.22 mmol) was added by syringe, and the mixture was stirred for 10 min. The lithium diisopropylamide solution was then cooled to –70 °C, and a solution of 2 g (7.5 mmol) diphenylmethylene glycine ethyl ester in 3 mL THF was slowly added. After 30 min stirring, 1.0 g (7.5 mmol) 3-bromobutyne was added over a 15 min period. The reaction mixture was allowed to warm to room

temperature and stirred for an additional 3 hours. The solvent was removed by evaporation and the product purified by column chromatography (1:1 CH_2Cl_2:hexane). The purified product was dissolved in 15 mL diethyl ether, and 15 mL of 1 M HCl was added. The mixture was stirred vigorously for 3 hours at room temperature. The two phases were separated, and the aqueous phase was washed twice with 10 mL diethyl ether. The aqueous layer was concentrated, and the residue was redissolved in 15 mL of 6 M HCl and heated to reflux for 6 hours. The water was removed, and the residue was taken up in 10 mL methanol and 5 mL propylene oxide and stirred for 12 hours at room temperature. The precipitate that formed was filtered and dried, yielding 0.25 g (27 %) of Y-Ile. The ^{1}H-NMR spectrum was in agreement with that reported previously [36]. Statistical mixtures of all 4 stereiosmers were obtained; the product was recrystallized four times from 5:1 isopropanol:water to give SS, RR-Y-Ile (d.e. = 81%) and SR, RS-Y-Ile (d.e. = 60%).

II.3.2 Purification of IleRS

E. coli strain MV1184 transformed with the multicopy plasmid pkS21[37] encoding the IleRS was kindly provided by Valerie de Crécy-Lagard and Paul Schimmel at the Scripps Research Institute. IleRS was expressed in 1 L (2xYT) cultures, and the collected protein was purified by ammonium sulfate fractionation, cold water dialysis, and anion exchange chromatography following an established protocol [38]. The concentration of IleRS stock was determined by the Bradford method to be 0.13 mM.

II.3.3 ATP/PP_i Exchange

ATP-PP_i exchange assays were performed as previously described [39] in 150 mL of reaction buffer (pH 7.6, 20 mM imidazole, 0.1 mM EDTA, 10 mM β-mercaptoethanol, 7 mM $MgCl_2$, 2 mM ATP, 0.1 mg/ml BSA, and 2 mM PP_i [^{32}P sodium pyrophosphate with a specific activity of 0.1 mCi/ml]) with 75 nM IleRS and concentrations of analogues from 10 mM to 1 mM. Aliquots of 15-20 mL of each reaction were removed at various time points and quenched in 0.5 ml of a solution of 200 mM PP_i, 7 % v/v $HClO_4$, and 3% w/v activated charcoal. The charcoal was washed twice with 0.5 ml of a solution of 10 mM PP_i with 0.5% v/v $HClO_4$ and resuspended in 0.5 ml of this solution. Each charcoal suspension was transferred to a 20-mL scintiallation vial, and 10 mL Safety-Solve liquid scintillation cocktail (Research Products Institute, Inc.) was added before counting on a Beckman Coulter liquid scintillation counter.

II.3.4 Protein expression

To test for analogue incorporation, a 50 mL culture of M9AA medium supplemented with ampicillin (200 mg/L), chloramphenicol (35 mg/L), 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.2 % glucose, and 1 mg/L thiamine was inoculated with a single colony of the isoleucine auxotrophic expression system AI-IQ[PQE15] [40]. After overnight growth at 37 °C, a 5 μL aliquot of culture was used to inoculate 50 mL of supplemented M9AA medium (for small-scale expressions) or a 1 mL aliquot was used to inoculate 1 L of medium (for large-scale expressions). When the culture reached an OD_{600} of 0.9-1.0, the cells were sedimented (5000*g*, 10 min, 4°C), washed twice with 0.9 % NaCl, and resuspended in 50 mL fresh supplemented M9 medium containing 19 natural amino acids (20 mg/L) but lacking isoleucine. For tests of incorporation, the cultures were divided

into aliquots, to which were added water (negative control), L-isoleucine (positive control), L-valine, SS-E-Ile, SR-E-Ile, SS-Y-Ile, or SR-Y-Ile. Tests of incorporation using once-recrystallized SS-E-Ile (d.e. = 68 %) were performed in 10 mL cultures; the concentrations of L-amino acid in the medium were 25, 50, and 125 mg/L. Studies of SS-E-Ile (d.e. = 95 %) and SR-E-Ile (d.e. = 98 %) were performed at 130 mg/L of the L-amino acid in 5 mL cultures, while studies of SS-Y-Ile (d.e. = 81%) and SR-Y-Ile (d.e. = 60%) were performed at 100 mg/L in 5 mL cultures. For large-scale expressions of protein samples for ^{1}H-NMR studies, the L-amino acid concentrations were 25 mg/L and 125 mg/L SS-E-Ile (d.e. = 76%) and 70 mg/L SR-E-Ile (d.e. = 91%). After 10 min of growth, mDHFR expression was induced by addition of 1 mM IPTG. After 4 h growth at 37 °C, the cells were sedimented (5000*g*, 10 min, 4°C), resuspended in 4 M urea, and frozen at –20 °C overnight. The cells were thawed, sonicated, and incubated for 30 min at 37 °C with 10 mg/mL DNase, 10 mg/mL RNase, and 10 mM $MgCl_2$ prior to purification. Protein expression was monitored by SDS-PAGE and Western blotting with antibodies recognizing the histidine tag of mDHFR (Qiagen, Inc., Santa Clarita, CA, USA) [41].

II.3.5 Mass spectrometry

Expressed mDHFR was purified from cell lysates by Ni affinity chromatography using a Ni-NTA Spin Kit (Qiagen). Purified mDHFR was incubated with trypsin in 50 mM ammonium bicarbonate buffer overnight at room temperature. The proteolysis product was purified by C18 ZipTip (Millipore), deposited in a matrix of α-cyano-4-

hydroxycinnamic acid, and analyzed by MALDI-TOF mass spectrometry on an Applied Biosystems Voyager DE Pro instrument.

II.3.6 ^{1}H-NMR spectroscopy

For ^{1}H-NMR, samples of purified protein were dissolved in 0.1% DCl in D_2O. Samples of SS-E-Ile (6 mg/L) were prepared in standard tubes (700 μL volume), while SR-E-Ile samples (4 mg/mL) were prepared in low volumes (300 μL) in solvent-matched tubes (Shigemi, Inc.). Spectra were collected on a Varian Inova NMR spectrometer with proton acquisition at 599.69 MHz and with water suppression by presaturation.

II.4 Results

II.4.1 Synthesis of unsaturated amino acids

Kazmaier has shown that *N*-benzyloxycarbonylglycine *trans*-crotyl ester undergoes a [3,3]-sigmatropic rearrangement [34] to yield Cbz-protected 2-amino-3-methyl-4-pentenoic acid (E-Ile). The reaction is stereoselective, giving mainly the SR, RS pair of stereoisomers (95%). Because it was of interest to compare the *in vivo* incorporation of the two diastereomers SS-E-Ile and SR-E-Ile, we applied the Kazmaier method also to Cbz-glycine *cis*-crotyl ester to obtain SS, RR-E-Ile (hereafter referred to as SS-E-Ile). The efficiency of rearrangement was lower than for the *trans*-ester, and the work-up requires an additional step because of a lower tendency of the SS, RR-E-Ile to crystallize, which contributed to a lower overall yield for the SS-analogue. The amino

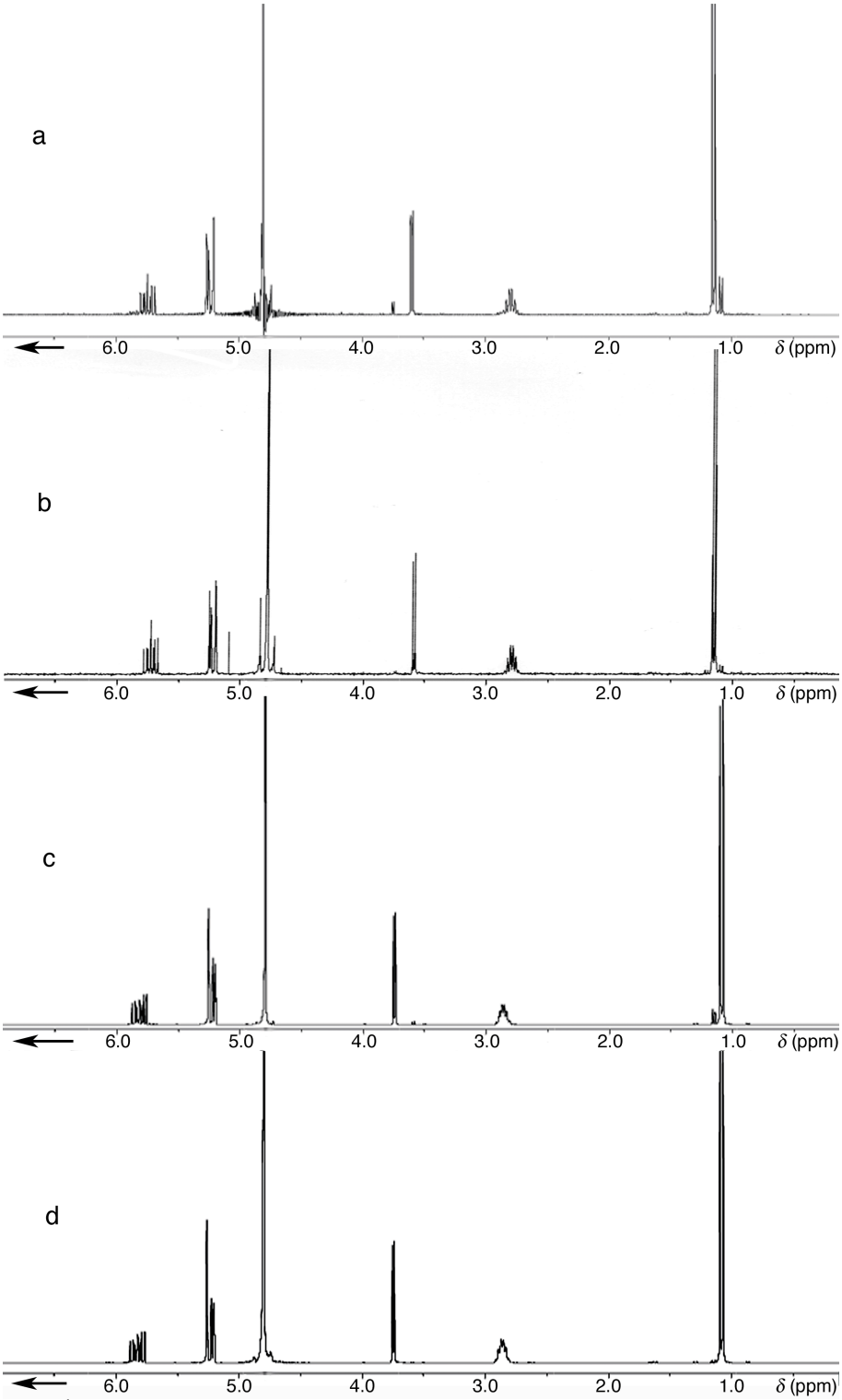


Figure II-5. ^{1}H-NMR spectra of a) SS-E-Ile after 1 recrystallization, b) SS-E-Ile after 6 recrystallizations, c) SR-E-Ile after 1 recrystallization, and d) SR-E-Ile after 6 recrystallizations.

acids were further purified by multiple recrystallizations to give SS-E-Ile (d.e. = 95 %) and SR-E-Ile (d.e. = 98 %) as determined by ^{1}H-NMR spectroscopy (Figure II-5).

The preparation of 2-amino-3-methyl-4-pentynoic acid was not stereoselective and gave a complex mixture of the SS, RR, SR, and RS stereoisomers. Recrystallization did not succeed in sufficiently separating SS, RR and SR, RS pairs; final products were ss-Y-Ile (d.e. = 81%) and ss-Y-Ile (d.e. = 60%) (Figure II-6).

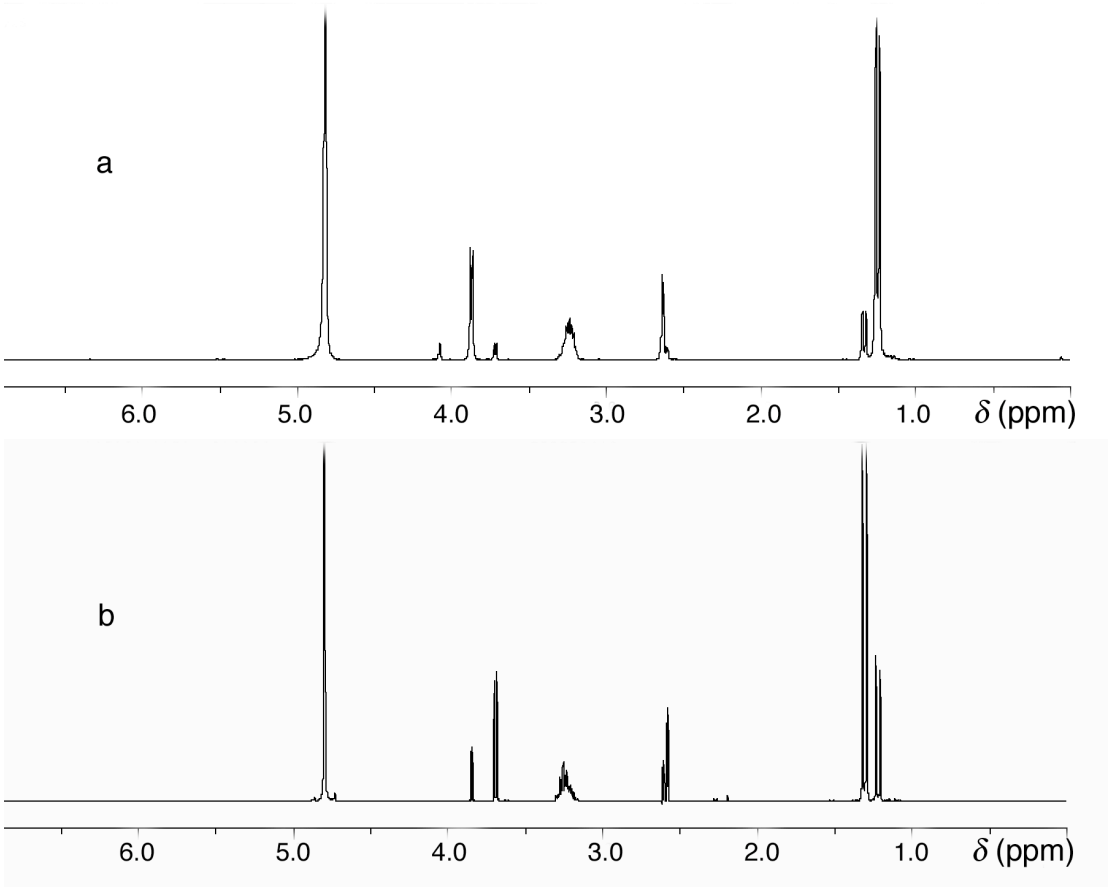


Figure II-6. ^{1}H-NMR spectra of a) SS-Y-Ile (d.e. = 81 %) and b) SR-Y-Ile (d.e. = 60%) after recrystallization.

II.4.2 Analogue incorporation

II.4.2.1 SS- and SR-Y-Ile

An *E. coli* strain rendered auxotrophic for isoleucine was used to assay the extent of *in vivo* incorporation of isoleucine analogues into mDHFR, a test protein readily expressed in bacterial cultures. SDS-PAGE and Western blotting of the total cellular protein produced in cultures supplemented with both SS-Y-Ile (d.e. = 81%) and SR-Y-Ile (d.e. = 60%) indicate expression of target protein (Figure II-7).

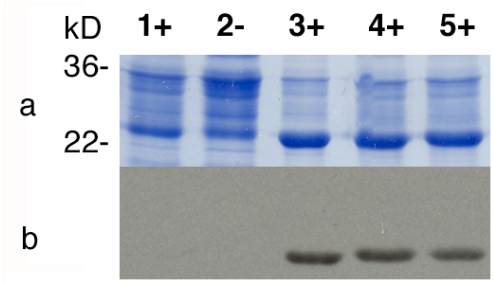


Figure II-7. a) SDS-PAGE and b) Western blot of proteins produced in Ile auxotrophic *E. coli* cultures supplemented with the 19 amino acids (lacking Ile) and 1) nothing, or 100 mg/L of 2) Ile, 3) Ile, 4) SR-Y-Ile, or 5) SS-Y-Ile. A + indicates induction of mDHFR expression; expression was not induced in lane 2. A similar amount of mDHFR is produced in cultures supplemented with SS-Y-Ile (d.e. = 81 %) and in those containing SR-Y-Ile (d.e. = 60%).

In the case of both SS- and SR-Y-Ile, MALDI-TOF mass spectra on tryptic fragments of mDHFR produced in medium supplemented with the analogue show peaks with the expected mass difference of -4 for each isoleucine in the peptide (Figure II-8). It is clear that we are able to incorporate at least one alkyne analogue into proteins *in vivo*, but without more careful purification of the stereoisomers it is impossible to say whether both or only one of these amino acids is translationally active.

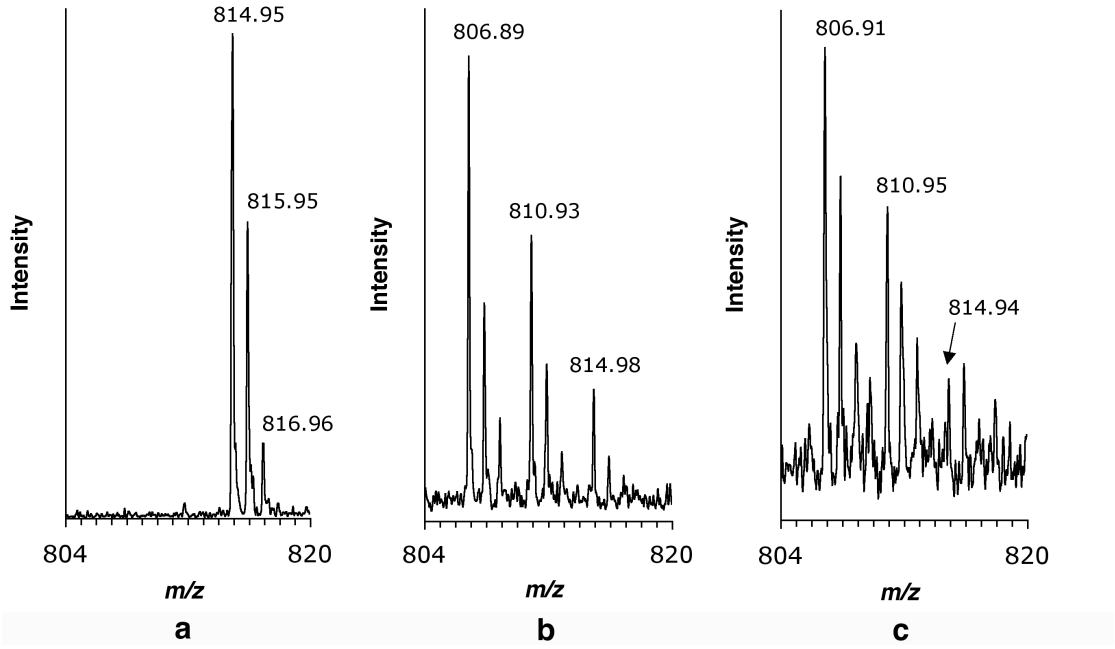


Figure II-8. MALDI-TOF spectra of tryptic fragments of mDHFR produced in medium supplemented with a) Ile, b) SS-Y-Ile (d.e. = 81 %), and c) SR-Y-Ile (d.e. = 60 %). Mass shifts of –4 per Ile residue in the peptide INIVLSR (residues 86 – 92, m/z = 814.5) in panels b) and c) indicates incorporation of an alkynyl analogue.

II.4.2.2 SS- and SR-E-Ile

In cultures supplemented with SS-E-Ile (d.e. = 95 %), SDS-PAGE and Western blotting again indicate that target protein is expressed in amounts comparable to the positive control (Ile); in contrast, cultures supplemented with SR-E-Ile (d.e. = 98 %) yielded significantly less protein (Figure II-9). This result is consistent with either a low level of translational activity of SR-E-Ile or with incorporation of residual SS-E-Ile in the amino acid sample. As discussed below, we believe the latter interpretation is correct.

MALDI-TOF mass spectra on tryptic fragments of mDHFR produced in medium supplemented with SS-E-Ile (d.e. = 95%) show signals shifted by the expected mass difference of -2 for each isoleucine residue in the peptide (Figure II-10). The MALDI

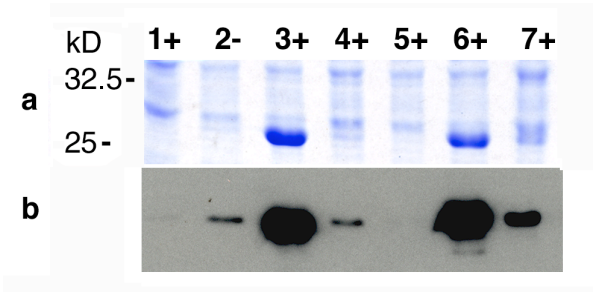


Figure II-9. a) SDS-PAGE and b) Western blot of proteins produced in Ile auxotrophic E. coli cultures supplemented with the 19 amino acids other than Ile and 1) nothing, or 130 mg/L of 2) Ile, 3) Ile, 4) norvaline, 5) Val, 6) SS-E-Ile, 7) SR-E-Ile. A + indicates induction of mDHFR expression; expression was not induced in lane 2. Significantly more mDHFR is produced in cultures supplemented with SS-E-Ile (d.e. = 95%) than in those containing SR-E-Ile (d.e. = 98%).

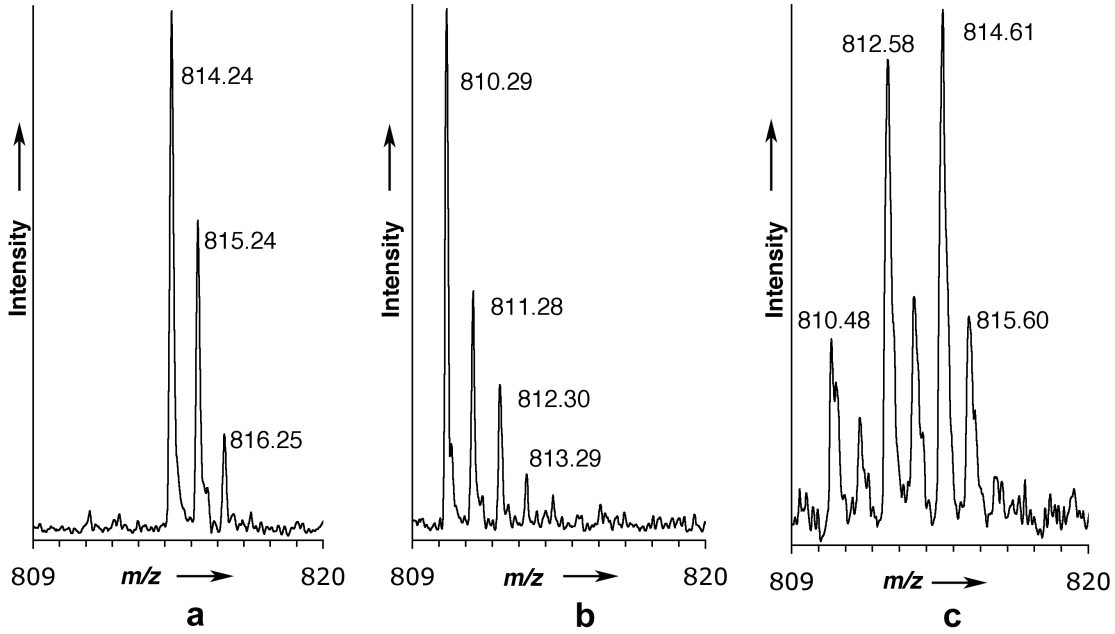


Figure II-10. MALDI-TOF spectra of tryptic fragments of mDHFR produced in medium supplemented with a) Ile, b) SS-E-Ile and c) SR-E-Ile. Mass shifts of –2 per Ile residue in the peptide INIVLSR (residues 86 – 92, m/z = 814.5) in panel b) indicate incorporation of SS-E-Ile. A lesser extent of incorporation of an unsaturated amino acid is also evident in proteins produced in culture supplemented with c) SR-E-Ile (d.e. = 98%).

spectra of fragments of target protein expressed in medium containing SR-E-Ile (d.e. = 98%) also show evidence of incorporation of an amino acid with a mass difference of -2. It is apparent that an unsaturated analogue did replace a fraction of isoleucine in each of these proteins, but of course it is not possible to distinguish between incorporation of SS-E-Ile and incorporation of SR-E-Ile from these data.

To determine the identity of the unsaturated amino acid that was incorporated into mDHFR in experiments with the SR-isomer, we purified protein from large-scale expressions conducted in medium supplemented with 70 mg/L SR-E-Ile (d.e. = 91%); protein yield was 4.1 mg/L. The alkene region of the ^{1}H-NMR spectrum of this protein (Figure II-11a) is identical to that of a protein expressed in medium supplemented with SS-E-Ile (d.e. = 76%) (Figure II-11b). In the amino acid spectra, the multiplet assigned to the internal alkene proton, H_i, is sensitive to stereochemistry; the chemical shift of this

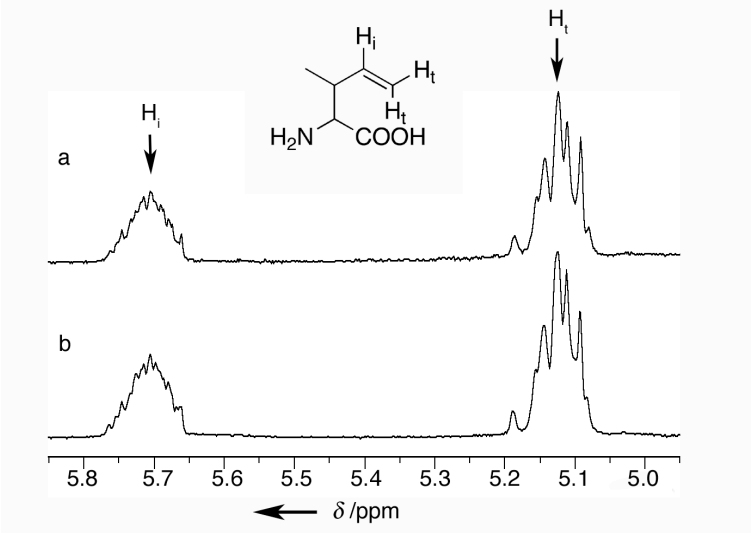


Figure II-11. ^{1}H-NMR spectra of mDHFR produced in medium supplemented with a) SS-E-Ile (d.e. = 76%) and b) SR-E-Ile (d.e. = 91%). Spectra of mDHFR containing only canonical amino acids show no peaks in this region.

proton differs by 0.1 ppm for the SS- and SR-isomers. We find it unlikely that this resonance would lose its sensitivity to stereochemistry in the protein context. We conclude that SR-E-Ile is a poor substrate for IleRS and that it is unable to compete with residual SS-isomer with respect to *in vivo* incorporation into proteins under the conditions used here.

To determine the extent of incorporation of ss-E-Ile, we prepared mDHFR in cultures supplemented with different levels of the analogue. SDS-PAGE and Western blotting of the total cellular protein produced in cultures supplemented with increasing concentrations of once-recrystallized SS-E-Ile (d.e. = 68 %) indicate increasing levels of target protein expression (Figure II-12). Integration of ^{1}H-NMR spectra of mDHFR produced in large-scale expressions in medium supplemented with 25 mg/L or 125 mg/L SS-E-Ile (d.e. = 76%) indicate levels of replacement of isoleucine by analogue of 63% and 72%, respectively. Protein yields are 12 and 22 mg/L, respectively.

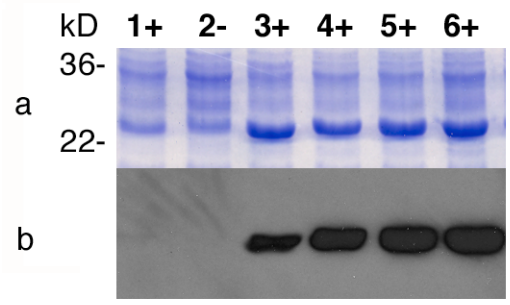


Figure II-12. a) SDS-PAGE and b) Western blot of proteins produced in Ile auxotrophic *E. coli* cultures supplemented with the 19 amino acids other than Ile and 1) nothing, 2) 25 mg/L Ile, 3) 25 mg/L Ile, 4) 25 mg/L ss-E-Ile, 5) 50 mg/L ss-E-Ile, and 6) 125 mg/L SS-E-Ile. A + indicates induction of mDHFR expression; expression was not induced in lane 2. mDHFR is produced in cultures supplemented with all concentrations of SS-E-Ile (d.e. = 68 %) in this range.

II.4.3 Kinetics of Activation of E-Ile by IleRS

The activation of the analogues by IleRS *in vitro* was investigated by ATP/PP_i exchange. Interestingly, SS-E-Ile was activated as fast as (or somewhat faster than) the natural substrate, Ile (k_{cat}/K_M = 0.30 μM^{-1} s^{-1}) [42], while SR-E-Ile was activated 40-fold more slowly than Ile (Table II-1). Because of the ~100-fold difference in the rates of activation of the SS- and SR-isomers, it not surprising that residual SS-E-Ile in the sample of the SR-isomer is responsible for the small amount of mDHFR produced in medium supplemented with SR-E-Ile (d.e. = 98 %).

Table II-1. Kinetics of Activation E-Ile by IleRS

analogue	V_{max} (M s^{-1})	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (μM^{-1} s^{-1})	Relative to Ile
SS-E-Ile	2.04 x 10^{-7}	32.9 ± 13	20.35 ± 0.23	0.756 ± 0.32	2.5
SR-E-Ile	4.54 x 10^{-8}	432 ± 101	4.53 ± 0.51	0.026 ± 0.0013	1/40
Ile	2.5 x 10^{-7}	58.2 ± 7.6	17.5 ± 4.5	0.3 ± 0.14	1

II.5 Discussion

Incorporation of amino acids analogues into proteins *in vivo* requires that the analogues (i) cross the cellular membrane; (ii) be charged to one or more tRNA(s); (iii) and be delivered to the growing end of the polypeptide chain. Our work indicates that none of these steps precludes efficient incorporation of at least one stereoisomer of both 2-amino-3-methyl-4-pentynoic acid (Y-Ile) and 2-amino-3-methyl-4-pentenoic acid (E-Ile). In the case of the alkene analogue, we show that the SS-isomer is greatly preferred

over the SR-isomer by the natural biosynthetic machinery of *E. coli*; indeed, we find no conclusive evidence of translational activity for SR-E-Ile.

Following cellular uptake and activation, the analogue must circumvent the editing pathways that normally limit misacylation of tRNAs. The selectivity (s) of an aaRS toward an amino acid is defined as the ratio of the rate of editing to the rate of activation [21]. The editing mechanism of *E. coli* isoleucyl-tRNA synthetase (IleRS) has been extensively studied [43, 44], and its selectivity for natural amino acids is high, ranging from s = 6000 for valine to s = 8.5×10^6 for alanine [21]. IleRS possesses two active sites: a synthetic site for binding of the amino acid prior to activation through formation of the aminoacyl adenylate and an editing site for removal of amino acids smaller than isoleucine (which fit into the editing pocket) [22, 23]. The SS-analogues tested in this study appears to circumvent the editing mechanism of IleRS, possibly because they are too large to fit into the editing site.

Our results show that IleRS is sensitive to stereochemistry at the β-carbon of E-Ile; only the SS-isomer of the isoleucine analogue is incorporated into protein at a measurable rate, and it is activated by the IleRS ~100-fold more rapidly than SR-E-Ile. This result is in agreement with previous binding studies that demonstrated that L-2-amino-3S-methylhexanoic acid binds preferentially to IleRS (K_a = 20 mM^{-1}); its diastereomer L-2-amino-3R-methylhexanoic acid binds to the enzyme with much a lower affinity (K_a = 0.6 mM^{-1}) [45]. It is also consistent with the fact that IleRS distinguishes L-isoleucine from L-allo-isoleucine [25, 27, 46]. Our data do not preclude the possibility, however, that discrimination between the *SS*- and *SR*-analogues occurs not in

the synthetic active site of IleRS but rather during some other translational step, such as editing by IleRS or binding to elongation factor-Tu or the ribosome.

Finally, the efficiency of substitution of SS-E-Ile and Y-Ile for Ile in recombinant proteins provides a simple and useful method for the incorporation of terminal double and triple bonds into proteins, giving the chemist access to versatile functional groups in proteins and protein-based materials.

II.6 References

1. van Hest JCM, Tirrell DA: **Protein-based materials, toward a new level of structural control**. *Chemical Communications* 2001:1897-1904.

2. Kiick KL, Saxon E, Tirrell DA, Bertozzi CR: **Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**:19-24.

3. Bae JH, Rubini M, Jung G, Wiegand G, Seifert MHJ, Azim MK, Kim JS, Zumbusch A, Holak TA, Moroder L, Huber R, Budisa N: **Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins**. *Journal of Molecular Biology* 2003, **328**:1071-1081.

4. Budisa N: **Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire**. *Angewandte Chemie-International Edition* 2004, **43**:6426-6463.

5. Bentin T, Hamzavi R, Salomonsson J, Roy H, Ibba M, Nielsen PE: **Photoreactive bicyclic amino acids as substrates for mutant Escherichia coli phenylalanyl-tRNA synthetases**. *Journal of Biological Chemistry* 2004, **279**:19839-19845.

6. Furter R: **Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in Escherichia coli**. *Protein Science* 1998, **7**:419-426.

7. Hendrickson TL, de Crecy-Lagard V, Schimmel P: **Incorporation of nonnatural amino acids into proteins**. *Annual Review of Biochemistry* 2004, **73**:147-176.

8. Kiick KL, Weberskirch R, Tirrell DA: **Identification of an expanded set of translationally active methionine analogues in Escherichia coli**. *Febs Letters* 2001, **502**:25-30.

9. Kim WY, George A, Evans M, Conticello VP: **Cotranslational incorporation of a structurally diverse series of proline analogues in an Escherichia coli expression system**. *Chembiochem* 2004, **5**:928-936.

10. Link AJ, Mock ML, Tirrell DA: **Non-canonical amino acids in protein engineering**. *Current Opinion in Biotechnology* 2003, **14**:603-609.

11. Wang L, Brock A, Herberich B, Schultz PG: **Expanding the genetic code of Escherichia coli**. *Science* 2001, **292**:498-500.

12. Wang L, Schultz PG: **Expanding the genetic code**. *Angewandte Chemie-International Edition* 2005, **44**:34-66.

13. Doring V, Mootz HD, Nangle LA, Hendrickson TL, de Crecy-Lagard V, Schimmel P, Marliere P: **Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway**. *Science* 2001, **292**:501-504.

14. Ibba M, Hennecke H: **Relaxing the substrate-specificity of an aminoacyl-transfer-rna synthetase allows in-vitro and in-vivo synthesis of proteins containing unnatural amino-acids**. *Febs Letters* 1995, **364**:272-275.

15. Datta D, Wang P, Carrico IS, Mayo SL, Tirrell DA: **A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins**. *Journal of the American Chemical Society* 2002, **124**:5652-5653.

16. Sharma N, Furter R, Kast P, Tirrell DA: **Efficient introduction of aryl bromide functionality into proteins in vivo**. *Febs Letters* 2000, **467**:37-40.

17. Tang Y, Tirrell DA: **Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo**. *Biochemistry* 2002, **41**:10635-10645.

18. Pauling L: **The Probability of Errors in the Process of Synthesis of Protein Molecules**. In *Festschrift Prof. Dr. Arthur Stoll*. Birkhauser Verlag; 1957:597-602.

19. Loftfield RB: **Frequency of Errors in Protein Biosynthesis**. *Biochemical Journal* 1963, **89**:82-&.

20. Jakubowski H, Fersht AR: **Alternative Pathways for Editing Non-Cognate Amino-Acids by Aminoacyl-Transfer Rna-Synthetases**. *Nucleic Acids Research* 1981, **9**:3105-3117.

21. Jakubowski H, Goldman E: **Editing of Errors in Selection of Amino Acids for Protein Synthesis**. *Microbiological Reviews* 1992, **56**:412-429.

22. Nureki O, Vassylyev DG, Tateno M, Shimada A, Nakama T, Fukai S, Konno M, Hendrickson TL, Schimmel P, Yokoyama S: **Enzyme structure with two catalytic sites for double-sieve selection of substrate**. *Science* 1998, **280**:578-582.

23. Silvian LF, Wang JM, Steitz TA: **Insights into editing from an Ile-tRNA synthetase structure with tRNA(Ile) and mupirocin**. *Science* 1999, **285**:1074-1077.

24. Schmidt E, Schimmel P: **Mutational Isolation of a Sieve for Editing in a Transfer-Rna Synthetase**. *Science* 1994, **264**:265-267.

25. Flossdorf J, Kula MR: **Ultracentrifuge Studies on Binding of Aliphatic Amino-Acids to Isoleucyl-Transfer-Rna Synthetase from Escherichia-Coli Mre 600**. *European Journal of Biochemistry* 1973, **36**:534-540.

26. Polet H, Conrad ME: **Influence of 3 Analogs of Isoleucine on in Vitro Growth and Protein Synthesis of Erythrocytic Forms of Plasmodium Knowlesi**. *Proceedings of the Society for Experimental Biology and Medicine* 1969, **130**:581-&.

27. Loftfield RB, Eigner EA: **Specificity of Enzymic Reactions . Aminoacyl-Soluble Rna Ligases**. *Biochimica Et Biophysica Acta* 1966, **130**:426-&.

28. Parker ED, Shive W, Skinner CG: **Biological Specificities of 4,5-Dehydro Analogues of Isoleucine and Alloisoleucine**. *Journal of Biological Chemistry* 1961, **236**:3267.

29. Clark TD, Ghadiri MR: **Supramolecular Design by Covalent Capture - Design of a Peptide Cylinder Via Hydrogen-Bond-Promoted Intermolecular Olefin Metathesis**. *Journal of the American Chemical Society* 1995, **117**:12364-12365.

30. Schafmeister CE, Po J, Verdine GL: **An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides**. *Journal of the American Chemical Society* 2000, **122**:5891-5892.

31. Nguyen ST, Grubbs RH, Ziller JW: **Syntheses and Activities of New Single-Component, Ruthenium- Based Olefin Metathesis Catalysts**. *Journal of the American Chemical Society* 1993, **115**:9858-9859.

32. Link AJ, Tirrell DA: **Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition**. *Journal of the American Chemical Society* 2003, **125**:11164-11165.

33. Hassner A, Alexanian V: **Synthetic Methods .12. Direct Room-Temperature Esterification of Carboxylic-Acids**. *Tetrahedron Letters* 1978:4475-4478.

34. Kazmaier U: **Synthesis of Unsaturated Amino-Acids by 3,3 -Sigmatropic Rearrangement of Chelate-Bridged Glycine Ester Enolates**. *Angewandte Chemie-International Edition in English* 1994, **33**:998-999.

35. Odonnell MJ, Polt RL: **A Mild and Efficient Route to Schiff-Base Derivatives of Amino-Acids**. *Journal of Organic Chemistry* 1982, **47**:2663-2666.

36. Aidene M, Barbot F, Miginiac L: **Reactivity of alpha-unsaturated organozinc compounds with N- (phenylsulfanyl)iminoesters - Application to the synthesis of monosubstituted or disubstituted unsaturated alpha-aminoacids**. *Journal of Organometallic Chemistry* 1997, **534**:117-127.

37. Shiba K, Schimmel P: **Functional Assembly of a Randomly Cleaved Protein**. *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**:1880-1884.

38. Shepard A, Shiba K, Schimmel P: **Rna-Binding Determinant in Some Class-I Transfer-Rna Synthetases Identified by Alignment-Guided Mutagenesis**. *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**:9964-9968.

39. Lawrence F, Blanquet S, Poiret M, Robertge.M, Waller J: **Mechanism of Action of Methionyl-Transfer Rna-Synthetase .3. Ion Requirements and Kinetic Parameters of Atp-Ppi Exchange and Methionine-Transfer Reactions Catalyzed by Native and Trypsin-Modified Enzymes**. *European Journal of Biochemistry* 1973, **36**:234-243.

40. Wang P, Tang Y, Tirrell DA: **Incorporation of trifluoroisoleucine into proteins in vivo**. *Journal of the American Chemical Society* 2003, **125**:6900-6906.

41. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual* edn 2nd. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 1989.

42. While the k_{cat} value measured for Ile (17 s^{-1}) is consistent with those reported in the literature [20], the K_M value (58 μM) was higher than reported values (4 – 5 μM) [20, 24].

43. Fersht A: *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. New York: W.H. Freeman & Company; 1999.

44. Fersht AR, Dingwall C: **Evidence for the Double-Sieve Editing Mechanism in Protein- Synthesis - Steric Exclusion of Isoleucine by Valyl-Transfer Rna-Synthetases**. *Biochemistry* 1979, **18**:2627-2631.

45. Flossdorf J, Pratorius HJ, Kula MR: **Influence of Side-Chain Structure of Aliphatic Amino-Acids on Binding to Isoleucyl-Transfer-Rna Synthetase from Escherichia- Coli Mre 600**. *European Journal of Biochemistry* 1976, **66**:147-155.

46. Loftfield TB, Hecht LI, Eigner EA: **Measurement of Amino Acid Specificity of Transfer Ribonucleic Acid**. *Biochimica Et Biophysica Acta* 1963, **72**:383-&.

C h a p t e r I I I

CONTROLLED DEGRADATION OF A CELL-ADHESIVE, ELASTOMERIC PROTEIN THROUGH INCORPORATION OF A FLUORINATED AMINO ACID*

III.1 Abstract

The design of biomaterials with controlled mechanical, cell-adhesive, and degradative properties is a common goal in tissue engineering and drug delivery systems. Towards this goal, a series of protein-based biomaterials were synthesized in an engineered bacterial production system. These modular proteins include domains from fibronectin that are known to adhere endothelial cells and elastin-derived repeating units to provide mechanical integrity. Fluorination of the protein by *in vivo* replacement of the amino acid isoleucine with the noncanonical amino acid 5,5,5-trifluoroisoleucine (5TFI) resulted in a tenfold decrease in degradation by the enzyme human leukocyte elastase compared to non-fluorinated protein. However, even after significant fluorination, the materials retain their ability to adhere endothelial cells in a sequence-specific manner. Incorporation of a noncanonical amino acid, without requiring a change in the encoding genetic sequence, represents a novel strategy to tune the rate of degradation of protein-based biomaterials without compromising cell adhesion.

**Manuscript prepared for submission by Sarah C. Heilshorn,* Marissa L. Mock,* and David A. Tirrell.*

**These authors contributed equally to this work.*

III.2 Introduction

Genetic engineering techniques allow the templated production of protein polymers with precisely controlled sequence, molecular weight, and functionality using microbial biosynthesis. Such techniques have been employed in the *de novo* design and synthesis of engineered proteins with a variety of novel physical and biological activities [1-8]. These materials have demonstrated potential in tissue engineering and reconstruction and drug delivery.

Elastin-like domains are of particular interest for engineered, protein-based biomaterials due to their high expression levels, ease of purification, biocompatibility, and tunable mechanical properties [9-14]. The extensive work of Urry and coworkers on the family of elastin-like polypentapeptides $(VPGZG)_x$, where Z is any amino acid, has shown that the hydrophobicity of the biopolymer can be used to tune the lower critical solution temperature (LCST) [15, 16]. Polymers are soluble at temperature below the LCST but phase separate into a polymer-rich coacervate as the temperature is increased. This LCST phase transition allows straightforward purification of elastin-like polymers after biosynthesis using a simple thermal cycling technique.

We have employed this method to purify a set of engineered proteins designed for use as implantable biomaterials. The protein sequences (Figure III-1) are a result of a modular design incorporating domains from fibronectin to adhere endothelial cells [10, 11], which are important for a healthy vasculature, and elastin-like domains to provide mechanical integrity [17, 18]. Sequence **CS5** contains the authentic cell-binding domain, which adheres human umbilical vein endothelial cells (HUVEC) in a sequence-specific

manner [11]. As a negative control, sequence **SC5** contains a scrambled cell-binding domain incapable of promoting sequence-specific HUVEC adhesion [19].

Proteins **CS5** (I=isolecuine) and **CS5-F** (I=5TFI):

M-MASMTGGQQMG-HHHHHHH-DDDDK-{LD-GEEIQIGHIPREDVDYHLYP-G[(VPGIG)$_2$VPGKG(VPGIG)$_2$]$_4$VP}$_3$-LE

T7 tag | His tag | Cleavage site | CS5 binding domain | Elastin-like domain

Proteins **SC5** (I=isolecuine) and **SC5-F** (I=5TFI):

M-MASMTGGQQMG-HHHHHHH-DDDDK-{LD-GEEIQIGHIPREVDDYHLYP-G[(VPGIG)$_2$VPGKG(VPGIG)$_2$]$_4$VP}$_3$-LE

T7 tag | His tag | Cleavage site | Scrambled CS5 binding domain | Elastin-like domain

Figure III-1. Amino acid sequences of the engineered proteins. Each protein has three cassettes of a cell-binding domain interspersed with an elastin-like domain. Protein **CS5** contains the authentic CS5 binding domain[11] while **SC5** contains a negative control, scrambled domain. Proteins **CS5-F** and **SC5-F** are identical to **CS5** and **SC5,** except they are synthesized in medium supplemented with 5TFI.

Depending on the specific medical application, e.g., drug delivery or tissue regeneration, the success of implanted biomaterials will depend on optimization of the *in vivo* degradation characteristics. Often, it is desirable to combine multiple degradation rates in one material. For example, a single system could combine rapid delivery of a pharmaceutical along with sustained release of growth factors for cell infiltration. Protein-based materials are degraded by a class of enzymes called proteases; however, native elastin is resistant to many of these proteases with the notable exception of elastase [20]. Human leukocyte elastase (HLE) is the predominant form of this protease that circulates the body in the blood stream [21]. HLE preferentially cuts after small, hydrophobic amino acids, and previous work in our laboratory showed that HLE prefers to cut after isoleucine in protein **CS5**. We hypothesized that we could alter the

degradation properties of this engineered protein by replacing isoleucine with a noncanonical amino acid.

The introduction of functional groups not contained within the 20 canonical amino acids into proteins is a valuable tool for the protein engineer, providing access to new chemical reactivity [22-26] and physical properties [27-31]. In many cases it is desirable to retain the biological activity of a protein upon introduction of these novel properties, requiring minimal disruption of the active, folded structure and making fluorinated amino acids of special interest. While fluorine is similar in size to hydrogen, the hydrophobicity of the CF_3 group is higher than the CH_3 group due to the low polarizability of fluorine [32], giving fluorination of proteins the potential to dramatically change physical properties of the protein without impairing its biological function [31, 33]. The ability to tune the rate of degradation of a biomaterial by varying the extent of incorporation of a noncanonical amino acid without requiring a change in the encoding genetic sequence would provide powerful control in target applications ranging from drug delivery to tissue engineering.

III.3 Methods

III.3.1 5,5,5-Trifluoroisoleucine (5TFI) synthesis and purification

5TFI was synthesized as previously described [33, 34] with minor modifications. Stereochemical purity was assessed by HPLC on a CROWNPAK CR (+) chiral column (Chiral Technologies, Inc.) with 1% perchloric acid/0.3% trifluoroacetic acid as the mobile phase. The crude product was a mixture of equal parts of the four stereoisomers

of 5TFI (D, L-, D-allo, L-allo-5TFI). It was recrystallized 3 times from 20% aqueous ethanol; in the final crystals only 1.7 % L-allo-5TFI remained.

The mixture of L- and D-5TFI was acetylated. D, L-5TFI (2.0 g, 0.013 mol) was suspended in 7 mL of 2 N NaOH and stirred in an ice bath. Another portion of 2 N NaOH (6 mL) and acetic anhydride (0.6 mL) were added. This addition was repeated 8 times at 2-minute intervals. The resulting clear solution was kept cold and stirred for 2 h before being neutralized with 18 mL of 6 N H_2SO_4. The solution was extracted with ether (3 x 50 mL), and the ether was dried over sodium sulfate and evaporated under vacuum. The yellow oil obtained was recyrstallized from 1:1 ethyl acetate:hexane, yielding 1.5 g white, crystalline N-acetyl-D, L-5TFI (74%).

N-acetyl-D, L-5TFI was enzymatically deacetylated to give pure L-5TFI. The acetylated amino acid (1.5 g, 0.0098 mol) was dissolved in 400 mL of 100 mM KH_2PO_4 (pH 7). Acylase I from porcine kidney (Sigma, 10 mg) was added, and the reaction was incubated at 37°C and followed by thin layer chromatography (4:1:1 n-butanol:acetic acid:water) with ninhydrin detection. When the concentration of free amino acid was no longer increasing (3 – 4 days), the reaction was acidified to pH 5 with concentrated HCl, filtered through a 0.22 μm filter, acidified further to pH 2, and extracted with ethyl acetate (3 x 75 mL). Evaporation of the ethyl acetate layer gave N-acetyl-D-5TFI. The aqueous layer was evaporated under vacuum, and the residue was taken up in methanol and filtered to remove a large portion of salt. The methanol was evaporated and the residue was dissolved in 100 mL of 0.1 N HCl and applied to an ion exchange column (Dowex 50WX4-100, Sigma). L-5TFI was eluted with 1 N NH_4OH and obtained as a white powder upon evaporation of the eluent under vacuum (150 mg, 10%). The final

product, by HPLC, is 2.7 % L-allo-5TFI and 3.2 % D-amino acid (a mixture of D-5TFI and D-allo-5TFI, which do not separate under the HPLC conditions employed).

III.3.2 Protein expression and purification

Proteins **CS5** and **SC5** were expressed as previously described [11]. To express proteins **CS5-F** and **SC5-F**, a competent isoleucine auxotrophic derivative of *E. coli* strain BL21(DE3), designated AI (*E. coli* B F^- *ompT hsdS*(r_B^- m_B^-)*gal dcm* λ(DE3) *ilvD691*), constructed in our laboratory [33] and harboring the plasmid pLysS (Qiagen), was transformed with, respectively, the plasmids pET28-CS5 and pET28-SC5 [17] to yield strains AI-pET28-CS5 and AI-pET28-SC5. To express proteins from these strains, a culture was grown overnight in 2xYT medium and used to inoculate 1 L of M9AA medium supplemented with the antibiotics chloramphenicol and kanamycin. At an OD_{600} of 0.8-1.0, the M9AA cultures were induced by adding 1 mM IPTG. After 20 additional minutes of growth, the cells were washed twice with 0.9% NaCl and resuspended in M9 medium containing 19 amino acids (excluding isoleucine) to a final volume of 1 L. The cultures were supplemented with 400 mg/L of 5TFI (effectively 100 mg/L of L-5TFI) and grown for 2 h. Fluorinated proteins were purified by Ni-affinity chromatography using Qiagen Ni-NTA agarose resin. Purity was assessed by SDS-PAGE and Western blotting with anti-T7 tag-horseradish peroxidase conjugate antibody (Amersham). Level of 5TFI incorporation was assessed by amino acid analysis at the University of California, Davis Molecular Structure Facility (Beckman 6300 amino acid analyzer).

3.2. Tryptic digest/MALDI

Purified proteins **CS5-F** and **SC5-F** were incubated with trypsin (50 mM ammonium bicarbonate buffer, overnight, room temperature). The proteolysis products were purified by eluting from a C18 ZipTip (Millipore) with 75:25 acetonitrile:0.1% trifluoroacetic acid, spotted on an analysis plate at 4°C, and analyzed by MALDI-TOF mass spectrometry on an Applied Biosystems Voyager DE Pro instrument.

III.3.3 LCST measurement

The LCST of proteins **CS5** and **CS5-F** was measured at 10 mg/ml in phosphate buffered saline (PBS), pH 7.4, by increasing the temperature at a rate of 30°C/h and measuring the percent transmission (measured in volts) at 300 nm on an Aviv model 62DS spectrophotometer (Lakewood, NJ).

III.3.4 Analysis of elastase degradation

For quantification of full-length chains, the degradation reaction was carried out at 37°C for 3 days in sodium borate buffer, pH 8, with 0.22 μM human leukocyte elastase (HLE, Elastin Products Company, Owensville, MO) and 100 μM protein. Samples were taken at 0, 1, 3, 6, 12, 24, 48, and 72 h and diluted with an equal amount of 2x SDS-sample buffer with β-mercaptoethanol and frozen at -20°C. Samples were boiled for 5 min, run on a 12% Tris-tricine gel at 150 V for 1 h, and transferred to poly(vinylidene fluoride) (PVDF) membrane for Western blot analysis using an anti-T7 tag-horseradish peroxidase conjugate antibody (Amersham) with a 10 second exposure. Densitometry was performed on Western blots using Image J (National Institutes of Health freeware image analysis program) to quantify the amount of whole-length protein remaining at

each time point. Control curves run in parallel indicated a linear relationship between protein concentration and Western blot intensity for both **CS5** and **CS5-F** (supporting information, Figure III-8).

For quantification of the number of cleaved peptide bonds, the degradation reaction was carried out at 37°C in sodium borate buffer, pH 8, with 0.22 μM HLE and 50 μM protein under constant mixing. The extent of reaction was characterized using 2,4,6-trinitrobenzene sulfonic acid at 4°C to quantify the number of N-termini in solution at 4 h.

III.3.5 Cell adhesion

Human umbilical vein endothelial cells (HUVEC, Bio Whittaker) were maintained in a 37°C, 5% CO_2 humidified environmental chamber. The cells were grown in Endothelial Growth Medium-2 (5% serum, Bio Whittaker), which was replaced every two days. Near confluent HUVEC cultures were passaged non-enzymatically by treatment with 0.61 mM EDTA (Gibco). Passages 2-5 were used.

Solutions of engineered proteins (1 mg/ml in PBS) and fibronectin (10 μg/ml in PBS) were adsorbed onto tissue culture polystyrene overnight at 4°C. Surfaces were rinsed with PBS, blocked with 0.2% heat-inactivated bovine serum albumin (BSA fraction V, Sigma) for 30 minutes, and rinsed.

HUVEC in suspension were labeled with a 5-μM solution of calcein acetoxymethyl ester (Molecular Probes) in serum-free Endothelial Basal Medium (EBM, Cell Applications, San Diego, CA) at room temperature for 30 min. Cells were rinsed twice and resuspended in EBM at 2.67×10^5 cell/ml. Cells (150 μl/well) were added to

adsorbed protein substrates in 96-well plates and incubated for 30 min. A solution of 21% w/w Percoll™ (Sigma) and PBS was added (200 μl/well) and plates were centrifuged at 100*g* for 10 min. Non-adherent cells were removed using harvesting frames (Molecular Devices) with the filters removed. PBS (100 μl/well) was added and fluorescence was measured using a Perkin Elmer HTS 7000 Bio Assay Reader at an excitation of 485 nm and emission of 538 nm. A cell adhesion index was calculated as the fluorescence reading of a test well divided by the fluorescence reading of HUVEC attached to fibronectin subjected to 1 *g*. The detachment force applied was estimated to be 26 pN using Archimedes' theorem [19].

III.4 Results and discussion

III.4.1 Protein synthesis and characterization

Using an engineered bacterial strain, the genetic message encoding the **CS5** protein can be alternatively read to produce protein **CS5-F**, a fluorinated version of **CS5**. The high isoleucine content of the **CS5** protein permits extensive fluorination through incorporation of the noncanonical amino acid 5TFI, which has demonstrated levels of isoleucine replacement from 85 – 93% in bacterial systems [33]. 5TFI was prepared as previously reported [33, 34] and used to express proteins **CS5-F** and **SC5-F** with yields of 0.83 mg/g and 0.71 mg/g wet cell mass, respectively. Nickel column purification yielded 5.4 mg pure **CS5-F** and 3.6 mg pure **SC5-F** from 1 L shake-flask fermentations. In contrast, proteins **CS5** and **SC5** express well and are easily purified using the thermal cycling technique to provide multi-gram quantities [17, 19]. Typical yields are 2-5 g pure

protein from 10 L batch fermentations. Further optimization is required to synthesize the fluorinated proteins above milligram yields.

To confirm the replacement of isoleucine with 5TFI, the engineered proteins were digested with the protease trypsin to yield protein fragments of predicted sequence, which were then analyzed by MALDI-TOF mass spectroscopy (Figure III-2). The peak at approximately 2576 Da has been assigned to two proteolytic fragments consisting of residues 136-161 and 262-286. These identical fragments contain five potential isoleucine replacement sites. Accordingly, the higher mass peaks are assigned to fragments with incorporation of one, two, three, four, and five 5TFI residues, each with a shift of 53.88 Da corresponding to the mass difference between isoleucine and 5TFI. Subsequent amino acid analysis reported 5TFI replacement of 82% of isoleucine residues

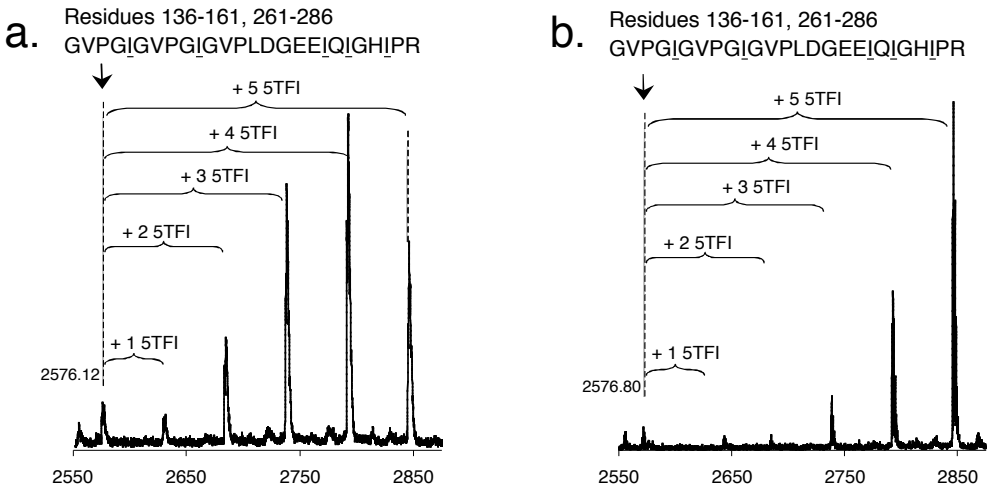


Figure III-2. MALDI-TOF of tryptic digest fragments of a) CS5-F and b) SC5-F. The expected mass of the unsubstituted peptide fragments comprising residues 136-161 and 261-286 = 2576. Peaks are apparent at masses expected for replacement of 1 through 5 isoleucines with 5TFI.

in **CS5-F** and 92% in **SC5-F**. This is consistent with results from previous studies demonstrating high incorporation efficiency of 5TFI into recombinant proteins [33].

We were interested in the effect of fluorination on the thermodynamic phase behavior of these proteins, as they are commonly purified through thermal cycling [7, 9]. As discussed above, proteins with elastin-like domains are known to exhibit an inverse temperature transition that is affected by the identity of the amino acid in the Z position [15, 16, 30, 35]. Relative to the most common pentapeptide repeat in bovine and porcine elastin, VPGVG [36, 37], the LCST is lowered when the amino acid occupying the Z position is more hydrophobic than valine and raised when Z is more hydrophilic. As expected, introducing the highly hydrophobic amino acid 5TFI into the Z position of the elastin-like domain results in a decrease of the LCST by more than 20°C (Figure III-3).

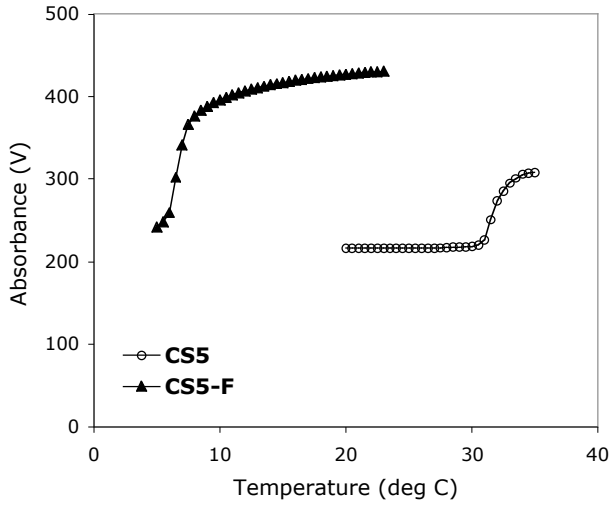


Figure III-3. Replacement of 82% of the isoleucine residues in CS5 with a fluorinated amino acid (CS5-F) decreases the LCST by 20°C, as evidenced by the turbidity of a 10 mg/ml in PBS solution, pH 7.4.

Therefore, the LCST of elastin-like proteins can be tuned by incorporation of fluorophilic 5TFI side chains, which may aid in optimization of thermal cycling purification techniques.

III.4.2 Protein degradation

Previous work in our laboratory has identified isoleucine as the favored HLE cut-site in these engineered, elastomeric proteins (S. C. Heilshorn, P. J. Nowatzki, T. Yamaoka, and D. A. Tirrell, manuscript in preparation). Therefore, incorporation of 5TFI into the isoleucine position was explored as a method to enhance HLE resistance. Due to the low expression levels of **CS5-F**, full kinetic analysis of the degradation reaction on the fluorinated protein was not possible. However, previous research in our laboratory has demonstrated that degradation of **CS5** follows traditional Michaelis-Menten kinetics with a catalytic constant, k_{cat}, of 0.033 s^{-1} and a Michaelis constant, K_m, of 2451 μM (supporting information, Figure 1II-7). Using these parameters, it is possible to predict the degradation rate of **CS5** and **CS5-F**, assuming that only the peptide bonds following isoleucine residues can be cleaved. The actual number of cleaved bonds in a reaction mixture can be quantified using 2,4,6-trinitrobenzene sulfonic acid to detect the concentration of N-termini. After 4 h of HLE degradation, 90% fewer peptide cleavages were observed on the fluorinated protein compared to the non-fluorinated protein (Figure III-4). These experimental results were in good agreement with the values predicted using the Michaelis-Menten model, which supports the assumption that HLE can only cleave peptide bonds following isoleucine in the fluorinated elastomer.

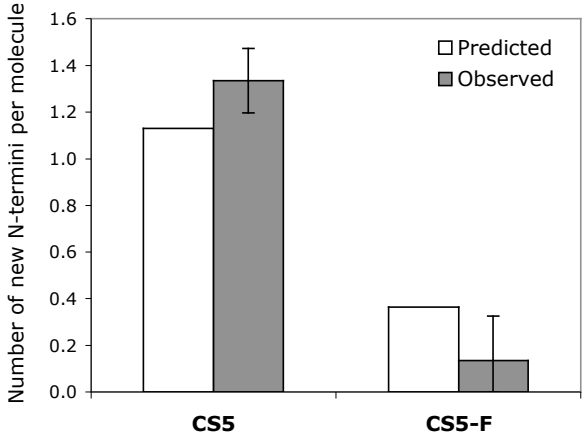


Figure III-4. HLE degradation of **CS5-F** produces 90% fewer new N-termini per original molecule than **CS5** after 4 h of reaction. Observed data represent two independent experiments, both testing three replicates of each substrate; error bars represent one standard deviation. Predicted data are based on the assumption that the reactions follow Michaelis-Menten kinetics and only the peptide bonds following isoleucine residues can be cleaved.

To examine the time course of degradation of **CS5** and **CS5-F**, Western analysis was used to determine the amount of full-length, intact protein remaining after HLE exposure for various times, and densitometry was employed to quantify the percent of full-length protein remaining at each time point (Figure III-5). Incorporation of 5TFI into the elastin-like protein significantly inhibited elastase activity. At 6 h, 46% of **CS5-F** remained intact, compared to 0% of **CS5**. Furthermore, full-length **CS5-F** was still detectable after 24 h exposure to HLE. Similar to the analysis performed above, the degradation reaction rates can be predicted using the Michaelis-Menten parameters for peptide cleavage after isoleucine residues. Using the simple assumption that each peptide cleavage results in the loss of one full-length protein chain, the predicted degradation rates for **CS5** and **CS5-F** are in good agreement with the observed values.

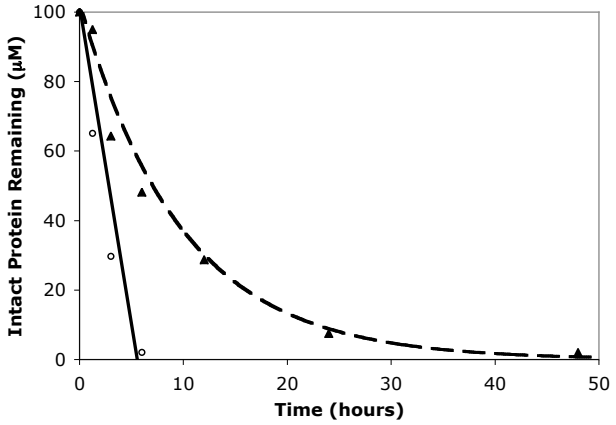


Figure III-5. HLE degradation of protein was monitored by Western blot and quantified using densitometry analysis of **CS5** (○) and **CS5-F** (▲) protein. No full-length **CS5** was observed after 6 h, while nearly half of **CS5-F** is still intact. The rate of chain degradation was predicted for **CS5** (dashed line) and **CS5-F** (solid line) using Michaelis-Menten parameters for the cleavage of peptide bonds after isoleucine residues.

III.4.3 Endothelial cell adhesion

We also investigated the ability of the fluorinated proteins **CS5-F** and **SC5-F** to promote cell adhesion. **CS5** was previously reported to be adherent to HUVEC in a sequence-specific manner [11]. Such adhesion is mediated through interactions with the REDV minimal binding sequence within the CS5 cell-binding domain. We wished to confirm that sequence-specific HUVEC adhesion would not be compromised by significant fluorination of the elastin-like regions. The ability of these engineered proteins to adhere HUVEC was examined using a buoyant centrifugation assay (Figure III-6). At a detachment force of 24 pN, protein **CS5-F** exhibited HUVEC adhesion greater than that of protein **CS5** and negative control proteins **SC5-F** and **SC5**, which contain scrambled cell-binding domains. Adhesion to fibronectin, which contains multiple cell-binding domains, was included as a positive control.

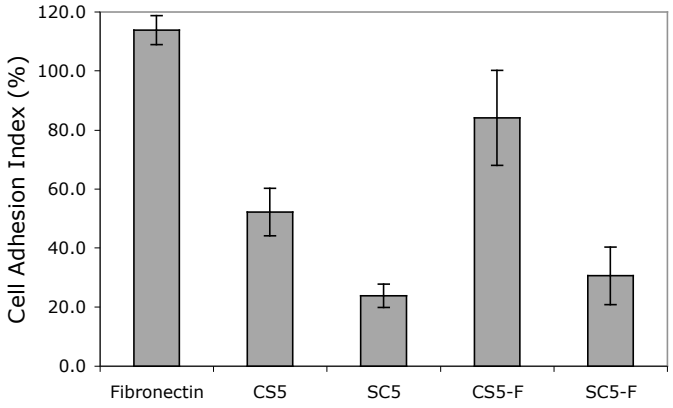


Figure III-6. Percent of HUVEC remaining adherent to adsorbed fibronectin and engineered proteins after 10 min exposure to 24 pN detachment force (100 *g*) relative to HUVEC remaining on adsorbed fibronectin after 10 min exposure to 0.24 pN (1*g*). Data represent three independent experiments in which six replicates of each substrate were tested; error bars represent one standard deviation.

These results suggest that 5TFI incorporation into these artificial proteins does not inhibit sequence-specific HUVEC binding. Therefore, fluorination of this engineered protein can successfully alter the thermodynamic behavior and proteolytic susceptibility without impairing the desired biological activity. While the observed increase in HUVEC adhesion to the fluorinated protein **CS5-F** relative to the non-fluorinated protein **CS5** is interesting, these results require further investigation. Ongoing research in our laboratory has shown similar context-dependence of adhesion strength to engineered proteins containing the CS5 domain.

We have demonstrated the ability to control the rate of proteolysis of elastin-like biomaterials through the incorporation of a noncanonical amino acid. From a single genetic message, two protein-based materials with varying degree of fluorination were created. Fluorination of the engineered elastomer retards elastase degradation of the

protein while also altering the thermodynamic phase behavior. The ability of the material to adhere endothelial cells through the sequence-specific interaction with the CS5 cell-binding domain is unaffected. Residue-specific incorporation of noncanonical amino acids into proteins is an additional tool for the biomedical engineer in the attempt to precisely control the material properties and biological activity of protein-based biomaterials.

III.5 Supporting information

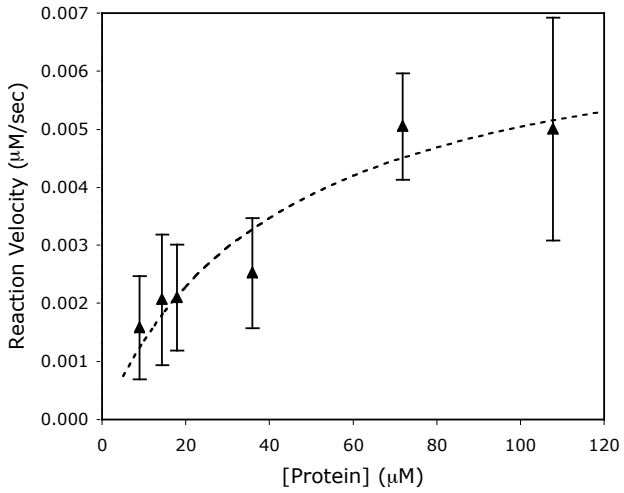


Figure III-7. Kinetic analysis of HLE degradation of **CS5**. Error bars represent a 90% confidence interval. The dashed line represents a best fit of the observed data to the Michaelis-Menten kinetic model.

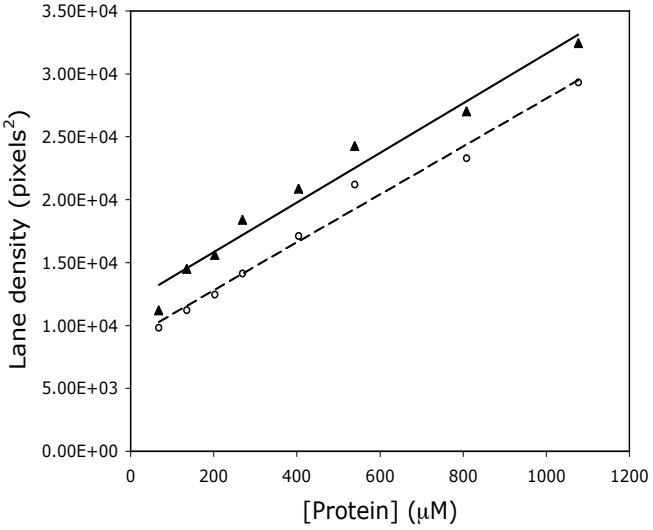


Figure III-8. Control curves for densitometry analysis of Western blots for **CS5** (○) and **CS5-F** (▲) with best-fit lines. Band intensity showed a linear increase with concentration up to 1 mM for both proteins.

III.6 References

1. McPherson DT, Xu J, Urry DW: **Product purification by reversible phase transition following Escherichia coli expression of genes encoding up to 251 repeats of the elastomeric pentapeptide GVGVP**. *Protein Expression and Purification* 1996, **7**:51-57.

2. Petka WA, Harden JL, McGrath KP, Wirtz D, Tirrell DA: **Reversible hydrogels from self-assembling artificial proteins**. *Science* 1998, **281**:389-392.

3. Krejchi MT, Atkins EDT, Waddon AJ, Fournier MJ, Mason TL, Tirrell DA: **Chemical Sequence Control of Beta-Sheet Assembly in Macromolecular Crystals of Periodic Polypeptides**. *Science* 1994, **265**:1427-1432.

4. Szela S, Avtges P, Valluzzi R, Winkler S, Wilson D, Kirschner D, Kaplan DL: **Reduction-oxidation control of beta-sheet assembly in genetically engineered silk**. *Biomacromolecules* 2000, **1**:534-542.

5. O'Brien JP, Fahnestock SR, Termonia Y, Gardner KCH: **Nylons from nature: Synthetic analogs to spider silk**. *Advanced Materials* 1998, **10**:1185-+.

6. Kaplan DL: **Spiderless spider webs**. *Nature Biotechnology* 2002, **20**:239-240.

7. Meyer DE, Chilkoti A: **Purification of recombinant proteins by fusion with thermally- responsive polypeptides**. *Nature Biotechnology* 1999, **17**:1112-1115.

8. Meyer DE, Chilkoti A: **Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: Examples from the elastin-like polypeptide system**. *Biomacromolecules* 2002, **3**:357-367.

9. Welsh ER, Tirrell DA: **Engineering the Extracellular Matrix: A Novel Approach to Polymeric Biomaterials. I. Control of the Physical Properties of Artificial Protein Matrices Designed to Support Adhesion of Vascular Endothelial Cells**. *Biomacromolecules* 2000, **1**:23-30.

10. Panitch A, Yamaoka T, Fournier MJ, Mason TL, Tirrell DA: **Design and biosynthesis of elastin-like artificial extracellular matrix proteins containing periodically spaced fibronectin CS5 domains**. *Macromolecules* 1999, **32**:1701-1703.

11. Heilshorn SC, DiZio KA, Welsh ER, Tirrell DA: **Endothelial cell adhesion to the fibronectin CS5 domain in artificial extracellular matrix proteins**. *Biomaterials* 2003, **24**:4245-4252.

12. Hong M, Isailovic D, McMillan RA, Conticello VP: **Structure of an elastin-mimetic polypeptide by solid-state NMR chemical shift analysis**. *Biopolymers* 2003, **70**:158-168.

13. Nagarsekar A, Crissman J, Crissman M, Ferrari F, Cappello J, Ghandehari H: **Genetic engineering of stimuli-sensitive silkelastin-like protein block copolymers**. *Biomacromolecules* 2003, **4**:602-607.

14. Betre H, Setton LA, Meyer DE, Chilkoti A: **Characterization of a genetically engineered elastin-like polypeptide for cartilaginous tissue repair**. *Biomacromolecules* 2002, **3**:910-916.

15. Urry DW, Luan C-H, Harris CM, Parker TM: **Protein-Based Materials with a Profound Range of Properties and Applications: The Elastin ΔTt Hydrophobic Paradigm**. In *Protein-Based Materials*. Birkahuser; 1997.

16. Urry DW, Gowda DC, Parker TM, Luan CH, Reid MC, Harris CM, Pattanaik A, Harris RD: **Hydrophobicity Scale for Proteins Based on Inverse Temperature Transitions**. *Biopolymers* 1992, **32**:1243-1250.

17. Di Zio K, Tirrell DA: **Mechanical properties of artificial protein matrices engineered for control of cell and tissue behavior**. *Macromolecules* 2003, **36**:1553-1558.

18. Nowatzki PJ, Tirrell DA: **Physical properties of artificial extracellular matrix protein films prepared by isocyanate crosslinking**. *Biomaterials* 2004, **25**:1261-1267.

19. Liu JC, Heilshorn SC, Tirrell DA: **Comparative cell response to artificial extracellular matrix proteins containing the RGD and CS5 cell-binding domains**. *Biomacromolecules* 2004, **5**:497-504.

20. Robert L, Robert AM, Jacotot B: **Elastin-elastase-atherosclerosis revisited**. *Atherosclerosis* 1998, **140**:281-295.

21. Oudijk EJD, Nieuwenhuis HK, Bos R, Fijnheer R: **Elastase mediated fibrinolysis in acute promyelocytic leukemia**. *Thrombosis and Haemostasis* 2000, **83**:906-908.

22. Zhang ZW, Wang L, Brock A, Schultz PG: **The selective incorporation of alkenes into proteins in Escherichia coli**. *Angewandte Chemie-International Edition* 2002, **41**:2840.

23. Liu HT, Wang L, Brock A, Wong CH, Schultz PG: **A method for the generation of glycoprotein mimetics**. *Journal of the American Chemical Society* 2003, **125**:1702-1703.

24. Link AJ, Tirrell DA: **Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition**. *Journal of the American Chemical Society* 2003, **125**:11164-11165.

25. Kiick KL, Saxon E, Tirrell DA, Bertozzi CR: **Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**:19-24.

26. Chin JW, Schultz PG: **In vivo photocrosslinking with unnatural amino acid mutagenesis**. *Chembiochem* 2002, **3**:1135-1137.

27. Bae JH, Rubini M, Jung G, Wiegand G, Seifert MHJ, Azim MK, Kim JS, Zumbusch A, Holak TA, Moroder L, Huber R, Budisa N: **Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins**. *Journal of Molecular Biology* 2003, **328**:1071-1081.

28. Bilgicer B, Xing X, Kumar K: **Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores**. *Journal of the American Chemical Society* 2001, **123**:11815-11816.

29. Budisa N, Rubini M, Bae JH, Weyher E, Wenger W, Golbik R, Huber R, Moroder L: **Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors**. *Angewandte Chemie-International Edition* 2002, **41**:4066-4069.

30. Tamura T, Yamaoka T, Kunugi S, Panitch A, Tirrell DA: **Effects of temperature and pressure on the aggregation properties of an engineered elastin model polypeptide in aqueous solution**. *Biomacromolecules* 2000, **1**:552-555.

31. Tang Y, Ghirlanda G, Vaidehi N, Kua J, Mainz DT, Goddard WA, DeGrado WF, Tirrell DA: **Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine**. *Biochemistry* 2001, **40**:2790-2796.

32. Bilgicer B, Kumar K: **Synthesis and thermodynamic characterization of self-sorting coiled coils**. *Tetrahedron* 2002, **58**:4105-4112.

33. Wang P, Tang Y, Tirrell DA: **Incorporation of trifluoroisoleucine into proteins in vivo**. *Journal of the American Chemical Society* 2003, **125**:6900-6906.

34. Muller N: **Synthesis of 5,5,5-Trifluoro-Dl-Isoleucine and 5,5,5-Trifluoro- Dl-Alloisoleucine**. *Journal of Fluorine Chemistry* 1987, **36**:163-170.

35. Urry DW: **Molecular Machines - How Motion and Other Functions of Living Organisms Can Result from Reversible Chemical-Changes**. *Angewandte Chemie-International Edition in English* 1993, **32**:819-841.

36. Yeh H, Ornstein-Goldstein N, Indik Z, Sheppard P, Anderson N, Rosenbloom JC, Cicila G, Yoon KG, Rosenbloom J: **Sequence Variation of Bovine Elastin Messenger-Rna Due to Alternative Splicing**. *Collagen and Related Research* 1987, **7**:235-247.

37. Sandberg LB, Leslie JG, Leach CT, Alvarez VL, Torres AR, Smith DW: **Elastin Covalent Structure as Determined by Solid-Phase Amino- Acid Sequencing**. *Pathologie Biologie* 1985, **33**:266-274.

C h a p t e r I V

LITHOGRAPHIC PATTERNING OF AN INTRINSICALLY PHOTOREACTIVE CELL-ADHESIVE PROTEIN*

IV.1 Abstract

This chapter describes a novel, simple method for the photolithographic patterning of cell-adhesive proteins. Intrinsically photoreactive proteins are synthesized in *Escherichia coli (E. coli)* through incorporation of the noncanonical photoactive amino acid *para*-azidophenylalanine. Upon ultraviolet irradiation at 365 nm, proteins form crosslinked films with elastic moduli that can be tuned by varying the concentration of photoreactive amino acid in the expression medium. Films of these proteins can be directly patterned using standard photolithographic techniques. Processing of irradiated films is performed under mild aqueous conditions, allowing these proteins to retain biological activity. We demonstrate the utility of this method of protein patterning by creating stable arrays of endothelial cells on an engineered protein "photoresist."

**Reproduced in part with permission from the Journal of the American Chemical Society, submitted for publication: Isaac S. Carrico, Sarah C. Heilshorn, Marissa L. Mock, Julie C. Liu, Paul J. Nowatzki, Stacey Maskarinec, Christian Franck, Guruswami Ravichandran, David A. Tirrell, Lithographic Patterning of Intrinsically Photoreactive Cell-Adhesive Proteins.*

IV.2 Introduction

Control of the spatial arrangement of proteins on surfaces is an essential factor in a number of emerging biotechnologies. Defining the location of specific proteins on the micro- or nanoscale improves the quality of protein microarrays, increases the sensitivity of biosensors [1, 2], and allows tissue engineering scaffolds to interact with multiple cell types [3]. Patterning is also a powerful tool in cell biology, where cell arrays are used to elucidate key factors that mediate migration, growth, and cell-cell interactions [4-6].

Although photolithography holds a preeminent place as a method to create patterns in the microelectronics industry, optical lithography of proteins has been hampered by the need either to use traditional chemical photoresists or to modify proteins chemically by attachment of photoactive functional groups; both methods can compromise protein function [7]. To circumvent these issues, new techniques such as microcontact printing and dip-pen nanolithography have emerged that allow direct placement of adhesive proteins or peptides on a surface [8, 9]. Despite these advances, the hallmarks of photolithography — high registry, throughput, and fidelity — are not yet matched by current methodologies.

Production of a protein "photoresist" without the need for post-translational chemical modification would require an intrinsically photoreactive protein. Recently, the incorporation of photoreactive noncanonical amino acids into proteins has been reported using both site-specific [10, 11] and residue-specific techniques [12]. Here we describe the microbial expression of artificial protein bearing the photosensitive noncanonical amino acid *para*-azidophenylalanine (*p*N_3Phe). Aryl azides have been previously shown

to mediate crosslinking efficiently in sensitive biological systems [13]. The recombinant proteins, designated artificial extracellular matrix proteins with aryl azides (aECM-N_3), belong to a family of engineered proteins designed to exhibit mechanical properties similar to those of native elastins [14], and to support adhesion of endothelial cells through a cell-binding domain derived from the CS5 region of fibronectin (Figure IV-1a) [15]. We demonstrate that these proteins can be efficiently crosslinked upon irradiation at 365 nm. The physical properties of the crosslinked films can be tuned by changing the extent of *p*N_3Phe incorporation, which is accomplished simply by changing the concentration of the noncanonical amino acid in the expression medium. Furthermore, thin films of such proteins can be patterned on surfaces using simple photolithographic techniques. We demonstrate the utility of the method by creating cell arrays through endothelial cell attachment to lithographically prepared protein patterns.

IV.3 Methods

IV.3.1 Protein expression

Samples of aECM-N_3 were expressed using a phenylalanine auxotrophic derivative of *E. coli* strain BL21(DE3), designated AF-IQ [16] and harboring the plasmid pNS-CS5-ELF [17], which encodes both the desired aECM sequence and the phes* gene for the A294G mutant *E. coli* PheRS. To express proteins from these strains, a culture was grown overnight in 2xYT medium and used to inoculate 1 L of M9AA medium supplemented with the antibiotics chloramphenicol and kanamycin. At an OD_{600} of 1.0, expression of target protein and T7 RNA polymerase was induced by adding 1 mM IPTG. After 10 additional minutes of growth, the cells were washed twice with 0.9%

NaCl and resuspended in M9 medium containing 19 amino acids (excluding phenylalanine) to a final volume of 1 L. The cultures were supplemented with either 25 mg/L phenylalanine (positive control) or up to 250 mg solid *p*N_3Phe and grown for 4 h. Protein expression was monitored by SDS-PAGE and Western blotting with anti-T7 tag-horseradish peroxidase conjugate antibody (Amersham).

IV.3.2 Protein purification

The aECM-N_3 protein was purified using a modified temperature cycling procedure previously described for elastin-containing aECM constructs [18, 19]. Expression cultures were pelleted at room temperature (10000*g*, 10 min, 25°C), resuspended in 20 mL of TEN buffer (10 mM Tris, 1 mM EDTA, 0.1 M NaCl) by sonication with a probe sonicator, and frozen at –20°C. To frozen lysate, 1 mM PMSF and 10 μg/mL each of DNase and RNase was added. This mixture was agitated for 4 h at 37°C and then centrifuged at room temperature (22000*g*, 60 min, 25°C), which is above the expected lower critical solution temperature of aECM-N_3. The target protein was extracted from the pellet by stirring overnight in 4 M urea at 4°C. This suspension was centrifuged (22000*g*, 60 min, 2°C), and the supernatant was dialyzed in 12-14 kD MWCO dialysis tubing against cold (4°C) distilled water for three days with six water changes. Precipitate formed during dialysis was removed by centrifugation (22000*g*, 60 min, 2°C). The supernatant, which contained aECM-N_3, was lyophillized.

IV.3.3 ^{1}H-NMR of aECM-N_3

Purified aECM-N_3 proteins were lyophilized completely and dissolved in DMSO-d_6 at 40 mg/mL. Spectra were collected on a Varian Inova NMR spectrometer with proton acquisition at 599.69 MHz

IV.3.4 FTIR spectroscopy

Infrared spectra were taken using a Perkin Elmer 1600 series FT-IR. Protein samples were drop-cast onto zinc selenide wafers from DMSO. To measure azide decomposition kinetics, a 10% aECM-N_3 solution in DMSO was spun onto a ZnSe crystal at 2000 rpm for 100 seconds. The resulting film was dried for 2 minutes at 50°C. Films were irradiated at 365 nm through at transparent mask in a Karl Suss mask aligner to mimic the protein lithography conditions. Infrared spectra were taken of the irradiated film at various time points until the azide peak was no longer changing.

IV.3.5 Mechanical testing of bulk films

Flims were prepared by drying aECM-N_3 solutions (10 wt % in DMSO) at 50°C overnight in Teflon molds followed by irradiation for 30 s approximately 8 inches from an unfiltered Oriel 100W medium pressure mercury lamp. The crosslinked samples were removed from the mold and immersed in 4°C water overnight to fully hydrate. Swollen flims were cut into testing strips of approximately 3 mm x 10 mm.

Tensile testing was performed using an Instron device with a constant strain rate of 10 % per minute on films equilibrated in a bath of PBS buffer at 37°C. Elastic

modulus (E) was obtained from the slope of the steepest part of the initial plot. Each protein sample was tested 3-6 times.

IV.3.6 Preparation of aminated glass surfaces

Glass coverslips were sonicated for 15 min in a saturated solution of potassium hydroxide in ethanol. Clean coverslips were rinsed under a stream of filtered (0.2 μm) doubly distilled water followed by a stream of ethanol, dried briefly with canned air, and then dried at 50°C for 30 min. Dried coverslips were immersed for 30 min in a freshly prepared solution of 1 mL 3-(trimethoxysilylpropyl)diethylenetriamine (DETA), 2.5 mL acetic acid, and 46.5 mL filtered water (2 % DETA in 5% acetic acid). Aminated coverslips were rinsed under a stream of water, rinsed under a stream of ethanol, dried with canned air, and cured for 4 h at 50°C.

IV.3.7 Preparation of PEG-modified slides

Aminated coverslips were placed in a covered dish containing a reservoir of pyridine. The coverslips were covered dropwise with a 100 mM solution of PEG-SPA-5000 (Nektar Therapeutics) in pyridine. After 12 h, the PEGylated coverslips were rinsed under a stream of water, rinsed under a stream of ethanol, dried with canned air, and used immediately.

IV.3.8 Spin coating of protein films

A 12.5 mg/mL solution of aECM-N_3 in DMSO was centrifuged for 1 min at 14000 rpm. This solution was added dropwise to cover the top of a PEGylated 12 mm circle glass coverslip (~8 μL). The coverslip was spun for 100 s at 1400 rpm on a

Specialty Coating Systems model P-6000 spin coater. Protein-coated slides were dried at 50°C for 30 min. Exposure of protein to sunlight was avoided until protein photolithography was complete.

IV.3.9 Protein lithography

Protein-coated slides were exposed for 30 s in a Karl Suss mask aligner (365 nm) under a chrome-on-quartz mask prepared by Dr. Michael Diehl at the California Institute of Technology by chrome deposition and stripping from a 3000 dpi transparency [20]. Irradiated slides were washed overnight in 0.05 % sodium dodecyl sulfate (SDS) to remove uncrosslinked protein from the masked regions and then rinsed for 6 h in doubly distilled, filtered water.

IV.3.10Cell culture

Human umbilical vein endothelial cells (HUVEC) were purchased from Clonetics and maintained in endothelial growth medium-2 (EGM-2, 2% serum, Clonetics, Walkersville, MD). Cells were kept in a humidified, 5% CO_2 environment at 37°C and passaged non-enzymatically using a 0.61 mM EDTA solution (Gibco, Grand Island, NY). Cells between passages 3 and 8 were used for all experiments.

All coverslips were sterilized by immersion in a 75% aqueous ethanol solution for at least 2 min and then dried using canned air. Substrates were placed in sterile 6-well polystyrene culture plates and blocked with a solution of 2% heat-denatured BSA in PBS for 30 min. For cell patterning studies, freshly harvested HUVEC cells were plated on the prepared substrates at a density of 4.2×10^4 cells/cm^2. Cell viability was measured

using a standard Wst-1 assay and cell adhesion was quantified using the buoyant centrifugation method described in Liu and coworkers [21].

IV.3.11 Phase contrast and fluorescence microscopy

Phase contrast pictures were taken on a Nikon Eclipse TE 300 microscope. Fluorescence pictures were taken on a Zeiss Axioplan II fluorescence microscope equipped with a monochrome Axiocam. To fix and fluorescently label cell patterns, the coverslips were placed in a 6-well plate, and each well was washed 3 times with 2 mL PBS before 1 mL ice-cold acetone was applied for exactly 1 min. The wells were again washed 3 times with 2 mL PBS before 2 mL of a 10% BSA solution was applied for 30 min at room temperature. After blocking, 0.2 μL of anti-T7 primary antibody (Novagen) was added and allowed to incubate at room temperature for at least 6 h. The wells were then washed three times with 2 mL PBS for 5 min without agitation. A secondary antibody/phalloidin solution composed of 862 μL PBS, 100 μL secondary antibody (Cy2-labeled anti-mouse, 0.5 mg/mL, Chemicon) and 38 μL rhodamine-phalloidin (Molecular Probes) was incubated with the samples in the dark for 1 h. Labeled samples were washed with 2 mL PBS for 10 minutes with agitation followed by 2 mL PBS for 5 minutes without agitation. The samples were then incubated with 1 mL of DAPI solution (0.3 μM in PBS) for 5 minutes at room temperature. Samples were rinsed 3 times with 2 mL PBS and mounted to a glass slide using filtered mounting solution of 1:1 PBS:glycerol and clear fingernail polish as sealant.

IV.3.12 Atomic force microscopy

Atomic force microscopy was conducted on dried, patterned aECM-N_3 protein films on PEGylated coverslips in constant-force, non-contact mode with an autoprobe M5 atomic force microscope (Parker Scientific).

IV.4 Results and Discussion

Large-scale production of the aECM-N_3 protein was accomplished through residue-specific incorporation of *p*N_3Phe in *E. coli*. This method of incorporation relies on competitive activation of phenylalanine and *p*N_3Phe by the phenylalanyl-tRNA synthetase, the enzyme responsible for charging phenylalanine to its cognate tRNA. The phenylalanyl-tRNA synthetase used for this study was a previously characterized mutant with relaxed substrate specificity [12, 22]. Proteins were expressed in a phenylalanine-auxotrophic *E. coli* strain grown in cultures supplemented with *p*N_3Phe and purified by taking advantage of the inverse temperature phase behavior of proteins with elastin-like repeats [23]. Incorporation efficiency was determined by integration of the aromatic proton signals in the ^{1}H NMR spectra of the purified proteins (Figure IV-1b); the extent of phenylalanine replacement varied from 13% to 53% depending on the concentration of *p*N_3Phe in the expression medium (Figure IV-2).

a.

MMASMTGGQQMGRKTHHHHHHHMG{LDGEEIQIGHIPREDVDYHLYPG[(VPGVG)$_2$(VPGFG)(VPGVG)$_2$]$_5$LP}$_3$LE

CS5 domain Elastin-like sequence

b.

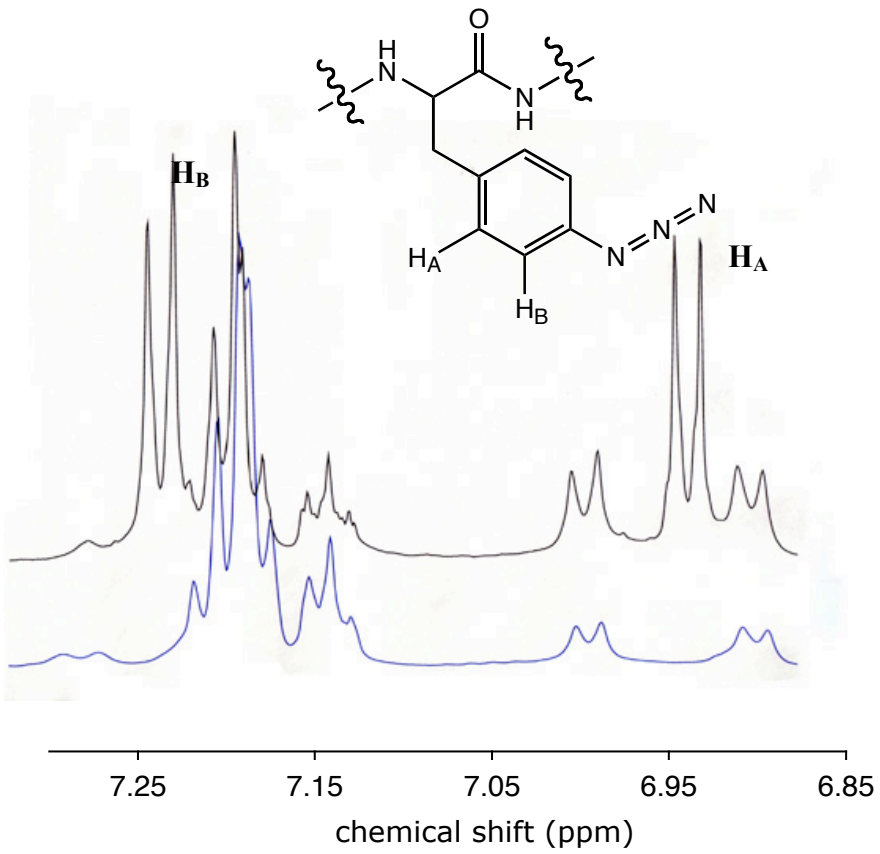


Figure IV-1. Design and production of aECM-N_3. (**a**) The aECM-N_3 primary sequence encodes a cell-binding domain (the CS5 region of fibronectin) and a structural domain (the pentapeptide VPGVG elastin-like repeat with periodic phenylalanine sites for incorporation of *p*N_3Phe). (**b**) ^{1}H NMR (600 MHz, DMSO-d_6) of the artificial protein expressed in the presence of phenylalanine (bottom spectrum) or in the presence of 250 mg/L *p*N_3Phe (top spectrum). Integration indicates 53% *p*N_3Phe incorporation.

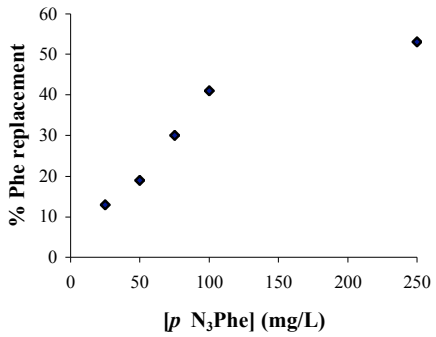


Figure IV-2. Incorporation of *p*N_3Phe into aECM-N_3 as a function of concentration in the expression medium.

Understanding the response of the designed photoreactive protein to irradiation is crucial for high-resolution pattern formation. We measured the rate of azide decomposition under irradiation by monitoring loss of the characteristic infrared (IR) asymmetric stretch at 2130 cm^{-1} (Figure IV-3a) [24]. Measurements were performed on thin films of aECM-N_3 spin-coated directly onto zinc selenide wafers and irradiated using a Karl Suss contact aligner filtered to 365 nm in constant intensity (7 mW/cm^2) mode, with a quartz wafer in place of the mask. Azide loss under these conditions was rapid, following first-order kinetics with a half-life of 34 seconds (Figure IV-3b). It is noteworthy that none of the other infrared bands were noticeably altered, implying that irradiation under the conditions used here specifically affects the aryl azide without substantial modification of any of the canonical amino acids. This is expected given that none of the canonical amino acids absorb above 310 nm [25].

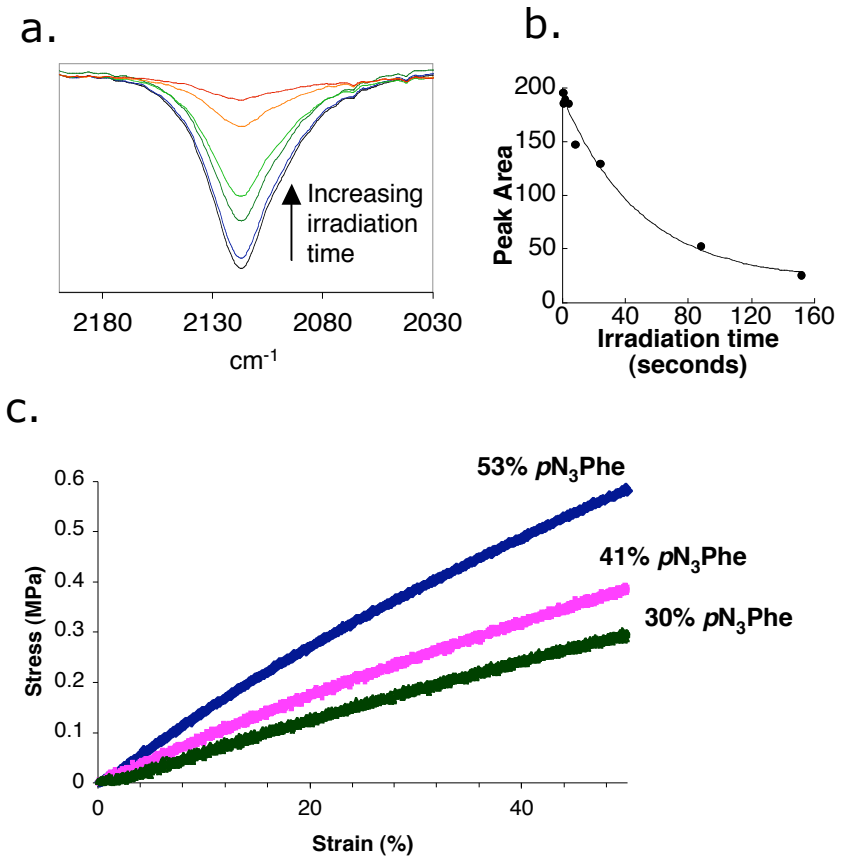


Figure IV-3. Characterization of azide decomposition by infrared spectroscopy and mechanical photocrosslinked films. **(a)** FT-IR spectrum of aECM-N_3 films as a function of irradiation time in the region of the aryl azide asymmetric stretch. Irradiation times are 0 (bottom), 0.5, 8, 24, 88, and 124 seconds (top). **(b)** Peak area versus irradiation time yields a first order exponential decay. **(c)** Uniaxial tensile testing of aECM-N_3 films with variable *p*N_3Phe incorporation.

Elastic moduli of irradiated aECM-N_3 bulk films were determined through uniaxial tensile testing under physiological conditions (Figure IV-3c). As expected, the elastic modulus correlated with the extent of *p*N_3Phe incorporation. Irradiated aECM-N_3 films in which 30, 41, and 53% of the encoded phenylalanine residues were replaced with *p*N_3Phe yielded elastic moduli of 0.53 ± 0.10, 0.94 ± 0.09, and 1.39 ± 0.09 MPa, which

are in the range of native elastin (0.3 – 0.6 MPa) [26]. Replacement of less than 20% of the encoded phenylalanine residues gave films that were too weak to test, and films made without *p*N_3Phe yielded no evidence of crosslinking. The fact that the modulus can be controlled simply by changing the *p*N_3Phe concentration in the expression medium is an attractive feature of the method, as recent work has highlighted the role of mechanical transduction mechanisms in mediating the physiology of adherent cells [27, 28].

To investigate the potential of our photoreactive proteins as substrates for studies of cell adhesion and growth, we created patterns of adherent endothelial cells on proteins patterned by photolithography. Protein films created by spin-coating 12.5% solutions of protein in dimethylsulfoxide directly onto prepared poly(ethylene glycol) (PEG)-coated glass coverslips were clear and homogeneous by optical microscopy. These protein films irradiated for 30 seconds at 365 nm through a chrome-on-quartz mask using a Karl Suss contact aligner. Stripping of the masked areas was accomplished by washing in mild aqueous detergent (0.05% aqueous sodium dodecasulfate).

Fluorescence immunolabeling with anti-T7-tag IgG antibody revealed that the aECM protein was localized only within the irradiated areas of the pattern (Figure IV-4a). Films prepared from protein lacking *p*N_3Phe formed no detectable patterns even after prolonged exposure times.

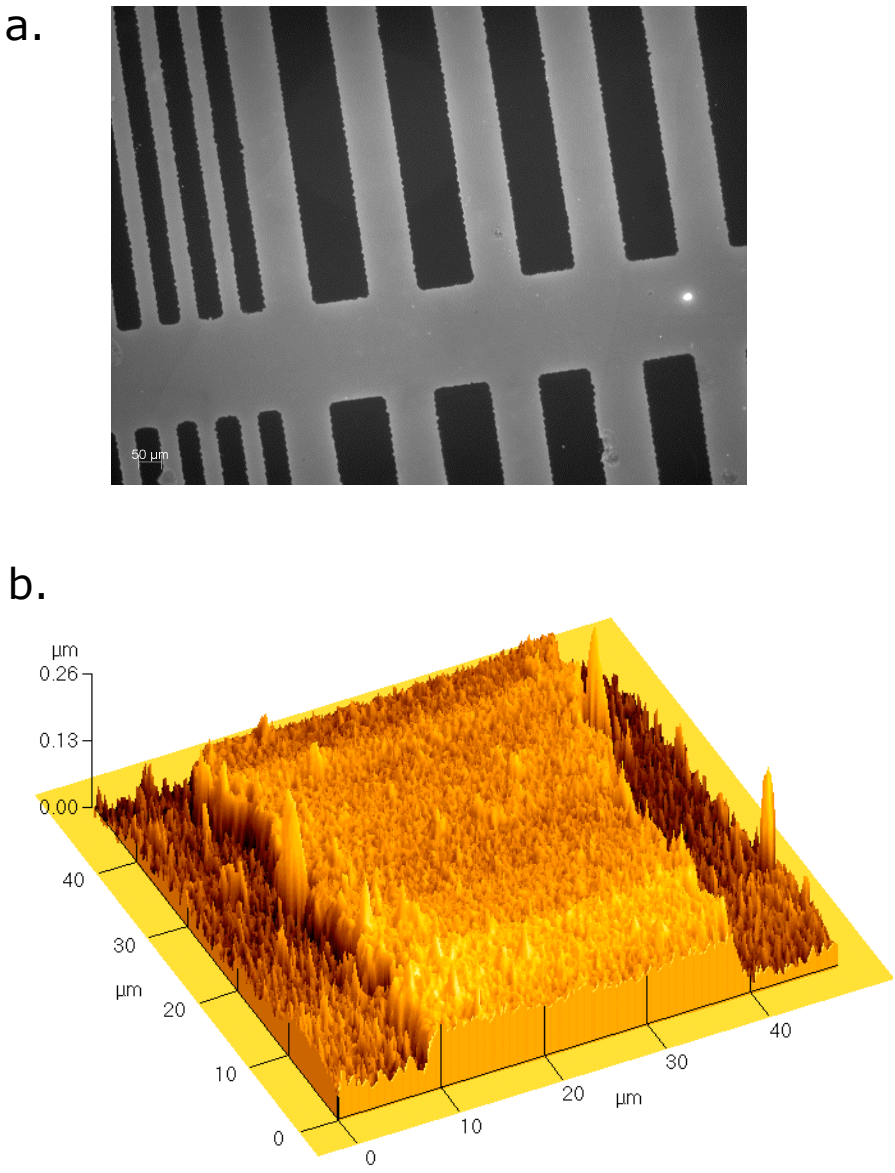


Figure IV-4. Characterization of photopatterned aECM-N_3 features by fluorescence and atomic force microscopy. **(a)** Fluorescence microscopy of photopatterned aECM-N_3. PEGylated glass slides were spin coated with aECM-N_3, irradiated for 30 seconds, and washed overnight in 0.05% SDS to produce well-defined protein features on a non-adhesive background. Protein patterns were immunolabeled with an anti-T7-tag primary antibody and an anti-mouse Cy2-conjugated secondary antibody. **(b)** AFM image of patterned aECM-N_3. The image was taken of a dried aECM-N_3 patterned film in constant-force, non-contact mode with an autoprobe M5 atomic force microscope.

The edge roughness of protein pattern features is a consequence of the roughness of the mask (created from a 3000 dpi print) and not the inherent resolution of the films (Figure IV-5). Non-contact AFM of dried aECM-N_3 patterns demonstrated uniform features, which varied in height depending upon the conditions used for spin coating. Films spun at 1400 rpm were 84 nm thick (Figure IV-4b), whereas those spun at 2000 rpm were approximately 4 nm thick. Protein patterns stored either dry or in aqueous solutions were stable for weeks.

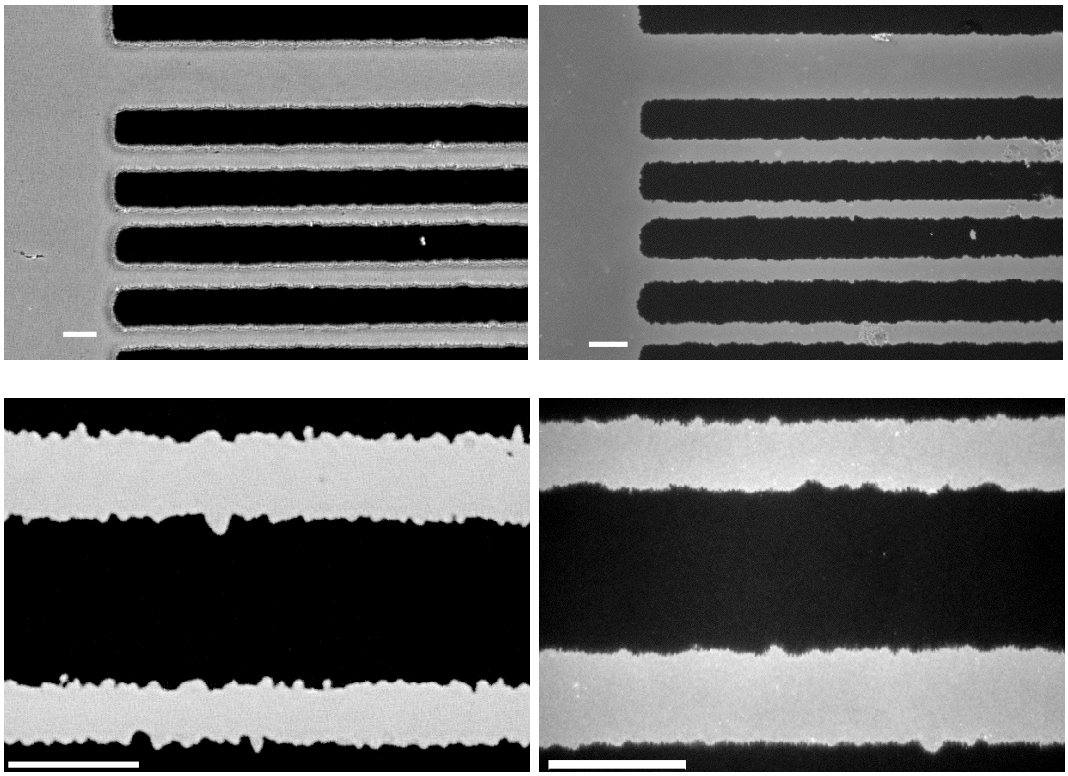
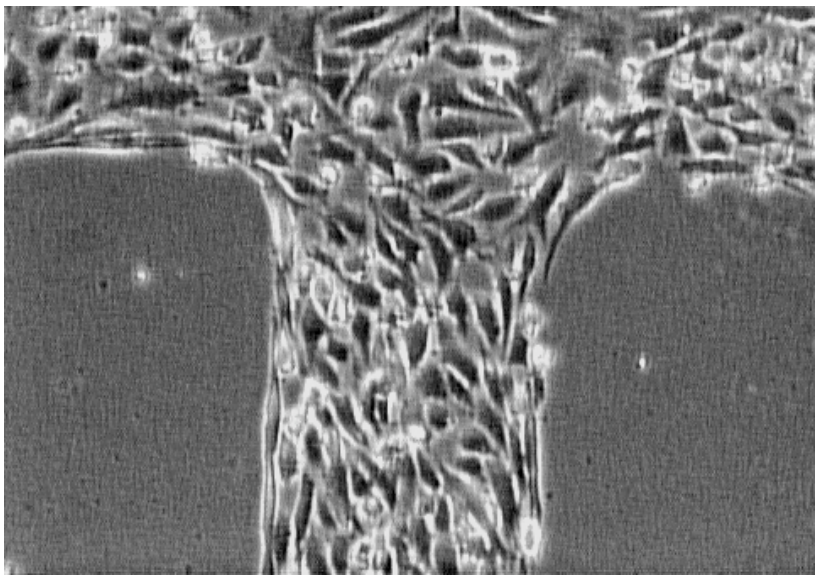


Figure IV-5. Phase contrast images of the chrome mask used in photopatterning (left panels) compared with fluorescence microscopy images of the protein patterns (right panels). Scale bars represent 50 microns. In the top panels (lower magnification), the protein pattern could be visually matched to precisely the region on the mask that created it. The bottom panels (higher magnification) show two separate regions with features of similar sizes.

To create cell arrays, HUVEC were plated on aECM-N_3 patterns in the absence of serum. After six hours of incubation, the medium was supplemented with 2% serum. After 24 hours of incubation, HUVEC exhibited a well-spread morphology (Figure IV-6a) and had proliferated to confluence exclusively within the patterned areas (Figure IV-6b). HUVEC monolayers in the interior of the patterned regions displayed morphology indistinguishable from monolayers grown on homogenous fibronectin coatings; however, HUVEC positioned along the aECM-N_3 pattern edges were elongated and oriented parallel to the pattern border (Figure IV-6a), consistent with previous studies. As expected, actin stress fibers within these elongated cells were aligned with the pattern edges (Figure IV-6b). Similar to other endothelial cell patterning techniques, this parallel cell alignment was generally observed for the first two to three cell layers adjacent to the pattern edges; therefore, cell alignment was enhanced on patterns with smaller feature sizes [29].

a.



b.

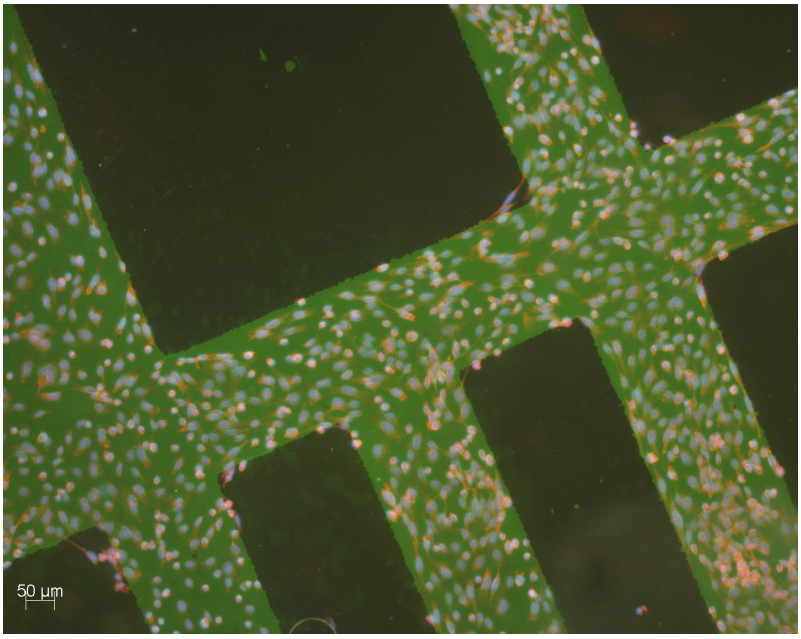


Figure IV-6. **(a)** Phase contrast and **(b)** fluorescence microscopy of HUVEC attached to photopatterned aECM-N_3. Cells were plated onto the prepared surfaces in serum free media and allowed to incubate 6 hours prior to supplementation with 2% serum. After 24 hours, the cells and substrates were fixed with acetone, immunolabeled with an anti-T7-tag primary antibody and anti-mouse Cy2-conjugated secondary antibody, and stained with phalloidin (specific to the actin cytoskeleton) and DAPI (specific to the nucleus).

In serum, HUVEC patterns were stable for 48 hours after reaching confluence, consistent with known behavior of PEG coatings as cell-resistant backgrounds [30]. At longer times, cells began growing beyond the protein pattern at the corners, probably in concert with synthesis and secretion extracellular matrix proteins.

The availability of intrinsically photoreactive proteins enables a facile new method for the patterning of proteins and cells. The technical simplicity of the method allows rapid production of samples with a variety of feature shapes and sizes, while permitting straightforward engineering of the elastic modulus of the crosslinked protein. The method represents a promising new approach to the study of adherent cells by providing exquisite control over mechanical properties, ligand-receptor interactions, and geometric shape. Applications in medical devices, tissue engineering, and array technologies are readily imagined.

IV.5 References

1. Templin MF, Stoll D, Schrenk M, Traub PC, Vohringer CF, Joos TO: **Protein microarray technology**. *Trends in Biotechnology* 2002, **20**:160-166.

2. Ekins RP: **Multi-Analyte Immunoassay**. *Journal of Pharmaceutical and Biomedical Analysis* 1989, **7**:155-168.

3. Tan W, Desai TA: **Layer-by-layer microfluidics for biomimetic three-dimensional structures**. In *Biomaterials*; 2004:1355-1364. vol 25.

4. Corey JM, Feldman EL: **Substrate patterning: an emerging technology for the study of neuronal behavior**. *Experimental Neurology* 2003, **184**:S89-S96.

5. Shim J, Bersano-Begey TF, Zhu XY, Tkaczyk AH, Linderman JJ, Takayama S: **Micro- and nanotechnologies for studying cellular function**. US Patent 2003.

6. Voldman J: **BioMEMS - Building with cells**. *Nature Materials* 2003, **2**:433-434.

7. Blawas AS, Reichert WM: **Protein patterning**. In *Biomaterials*; 1998:595-609. vol 19.

8. Whitesides GM, Ostuni E, Takayama S, Jiang XY, Ingber DE: **Soft lithography in biology and biochemistry**. In *Annual Review of Biomedical Engineering*: 2001:335-373.

9. Lee KB, Park SJ, Mirkin CA, Smith JC, Mrksich M: **Protein nanoarrays generated by dip-pen nanolithography**. In *Science*; 2002:1702-1705. vol 295.

10. Chin JW, Martin AB, King DS, Wang L, Schultz PG: **Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli**. In *Proceedings of the National Academy of Sciences of the United States of America*; 2002:11020-11024. vol 99.

11. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG: **Addition of p-azido-L-phenylaianine to the genetic code of Escherichia coli**. In *Journal of the American Chemical Society*; 2002:9026-9027. vol 124.

12. Kirshenbaum K, Carrico IS, Tirrell DA: **Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues**: 2002.

13. Pendergrast PS, Chen Y, Ebright YW, Ebright RH: **Determination of the Orientation of a DNA-Binding Motif in a Protein DNA Complex by Photo-**

Cross-Linking. *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**:10287-10291.

14. DiZio KA, Tirrell DA: **Mechanical properties of artificial protein matrices engineered for control of cell and tissue behavior**. In *Macromolecules*; 2003:1553-1558. vol 36.

15. Heilshorn SC, DiZio KA, Welsh ER, Tirrell DA: **Endothelial cell adhesion to the fibronectin CS5 domain in artificial extracellular matrix proteins**. In *Biomaterials*; 2003:4245-4252. vol 24.

16. Yoshikawa E, Fournier MJ, Mason TL, Tirrell DA: **Genetically-Engineered Fluoropolymers - Synthesis of Repetitive Polypeptides Containing P-Fluorophenylalalanine Residues**. *Macromolecules* 1994, **27**:5471-5475.

17. Sharma N: **Biosynthetic introduction of aryl bromide functionality into proteins [Doctoral]**. Amherst: University of Massachusetts: 2001.

18. Di Zio K, Tirrell DA: **Mechanical properties of artificial protein matrices engineered for control of cell and tissue behavior**. *Macromolecules* 2003, **36**:1553-1558.

19. Nowatzki PJ, Tirrell DA: **Physical properties of artificial extracellular matrix protein films prepared by isocyanate crosslinking**. *Biomaterials* 2004, **25**:1261-1267.

20. Madou M: *Fundamentals of Microfabrication* edn 2. New York: CRS Press; 2001.

21. Liu JC, Heilshorn SC, Tirrell DA: **Comparative cell response to artificial extracellular matrix proteins containing the RGD and CS5 cell-binding domains**. *Biomacromolecules* 2004, **5**:497-504.

22. Ibba M, Soll D: **Aminoacyl-tRNA synthesis**. In *Annual Review of Biochemistry*; 2000:617-650. vol 69.

23. Welsh ER, Tirrell DA: **Engineering the extracellular matrix: a novel approach to polymeric biomaterials. I. Control of the physical properties of artificial protin matrices designed to support adhesion of vascular endothelial cells**. In *Biomacromolecules*; 2000:23-20. vol 1.

24. Pretsch E, Simon W, Seibl J, Clerc T: **Infrared Spectroscopy: Compounds of the type X=Y=Z**. In *Spectral Data for Structure Determination of Organic Compounds*. Springer-Verlag; 1989:I80.

25. Wetlaufer DB: **Ultraviolet Spectra of Proteins and Amino Acids**. In *Advances in Protein Chemistry*. Academic Press; 1962:303-390. vol 17.

26. Fung YC: *Biomechanics: Mechanical Properties of Living Tissues* edn 2nd. New York: Springer-Verlag; 1993.

27. Ingber DE: *Mechanosensation through integrins: Cells act locally but think globally*, vol 100; 2003.

28. Engler A, Bacakova L, Newman C, Hategan A, Griffin M, Discher D: **Substrate compliance versus ligand density in cell on gel responses**. In *Biophysical Journal*; 2004:388a-388a. vol 86.

29. Zhang SG, Yan L, Altman M, Lassle M, Nugent H, Frankel F, Lauffenburger DA, Whitesides GM, Rich A: **Biological surface engineering: a simple system for cell pattern formation**. In *Biomaterials*; 1999:1213-1220. vol 20.

30. Mrksich M, Dike LE, Tien J, Ingber DE, Whitesides GM: **Using microcontact printing to pattern the attachment of mammalian cells to self-assembled monolayers of alkanethiolates on transparent films of gold and silver**. In *Experimental Cell Research*; 1997:305-313. vol 235.

Chapter V

PROGRESS TOWARD CLONING THE GENE OF AN RGD-CONTAINING PROTEIN FOR PHOTOPATTERNING*

V.1 Abstract

We have previously described a new method for the production of cell-adhesive photolithographic protein patterns (Chapter 4) using the noncanonical amino acid *p*-azidophenylalanine (*p*N_3Phe). We demonstrated the procedure using an artificial extracellular matrix (aECM) protein [1] containing the CS5 cell-binding sequence derived from the IIICS region of fibronectin that is known to selectively adhere endothelial cells [2]. Previous work in our group has shown that endothelial cells attach both more quickly and strongly to an aECM protein containing an RGD-based sequence [3] from the CCBD of fibronectin [4]. We wished to create an expression strain harboring a plasmid encoding the artificial gene for an RGD-based sequence with phenylalanine residues within the elastin-like domain, which would permit photopatterning through the incorporation of *p*N_3Phe. The cloning strategy and progress toward the final construct will be described in this chapter.

**This work was performed with Stacey Maskarinec.*

V.2 Introduction

As described in Chapter 4, we have recently demonstrated that incorporation of the noncanonical amino acid *p*-azidophenylalanine (*p*N3Phe) into a protein, dubbed artificial extracellular matrix protein with aryl azides (aECM-N_3), permits patterning of the protein on a surface using traditional photolithographic techniques. Further, we used this technique to create cell patterns through selective attachment of human umbilical vein endothelial cells to the aECM-N_3 protein, which contains a known endothelial cell-binding domain, when patterned on a non-adhesive background of poly(ethylene glycol) (PEG).

The sequence of the aECM-F protein we studied (Figure V-1) contained elastin-like repeats to provide structural integrity, regular Phe residues within the elastin-like regions as sites of crosslinking, and repeats of a cell-adhesive sequence (CS5) derived from the IIICS region of fibronectin that is known to adhere endothelial cells but not

CS5-F: MMASMTGGQQMGHHHHHHMG(LD-GEEIQIGHIPREDVDYHLY-PG((VPGVG)$_2$VPG**F**G(VPGVG)$_2$)$_5$VP)$_3$-LE

SC5-F: MASMTGGQQMGHHHHHHMKL{LDASFLD-GEEIQIGHIPREVDDYHLY-ASA[(VPGVG)$_2$VPG**F**G(VPGVG)$_2$]$_5$VP}$_3$LE

RGD-F: MASMTGGQQMGHHHHHHMKL{LDASFLD-YAVTGRGDSPASSKPIA-ASA[(VPGVG)$_2$VPG**F**G(VPGVG)$_2$]$_5$VP}$_3$LE

RDG-F: MASMTGGQQMGHHHHHHMKL{LDASFLD-YAVTGRDGSPASSKPIA-ASA[(VPGVG)$_2$VPG**F**G(VPGVG)$_2$]$_5$VP}$_3$LE

Figure V-I. Amino acid sequences of the artificial extracellular matrix proteins. The CS5-F protein was previously cloned by Nandita Sharma [1] and is discussed in Chapter 4. The target proteins for cloning are a construct containing a different cell-binding domain (RGD-F) and the negative control proteins for both cell-binding domains (RDG-F and SC5-F).

fibroblasts, platelets, or smooth muscle cells [2]. We will refer to this construct as CS5-F. Previous work in our laboratory on similar constructs (lacking Phe residues) [3] showed that endothelial cells form stronger attachments and form them more quickly when aECM proteins with repeats of an RGD-containing sequence derived from the CCBD region of fibronectin [4] are used. We wished to create an expression strain harboring a plasmid copy of the artificial gene for the RGD-containing version of the aECM protein we used previously. We also undertook to clone the appropriate negative control proteins for both the CS5-F and RGD-F sequences; these are designated SC5-F and RDG-F, respectively. The set of proteins is generically designated CBD-F for cell-binding domain construct with Phe residues. The sequences of the CBD-F (Figure V-1) proteins were designed based on CS5-F to contain repeats of the appropriate cell-binding domains as well as regular Phe residues within the elastin-like regions to permit photocrosslinking.

V.2.1 Cloning strategy

The cloning strategy to obtain this sequence (Figure V-2) was based on one developed by Liu and Tirrell [5].

Synthetic DNA oligonucleotides encoding the desired CBDs (Figure V-3) were designed and ordered commercially. Phosphorylated oligonucleotides were ligated into the pUC19 vector at the *Eco*R I / *Bam*H I sites. The cell-binding domain regions were cut out of the pUC19-SC5, pUC19-RGD, or pUC19-RDG vectors with *Eco*R I / *Bam*H I

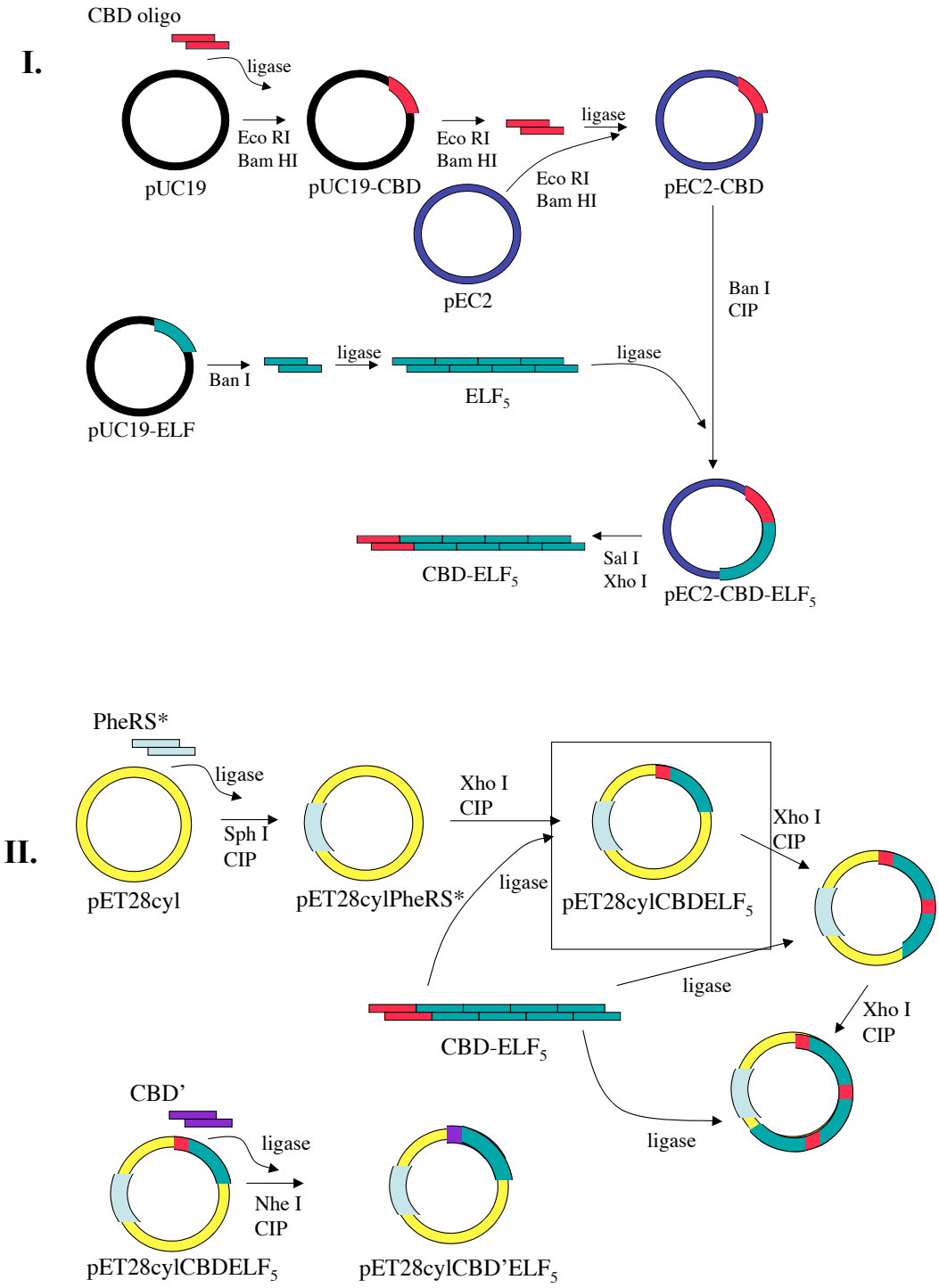


Figure V-2. The cloning strategy for CBD-F proteins.

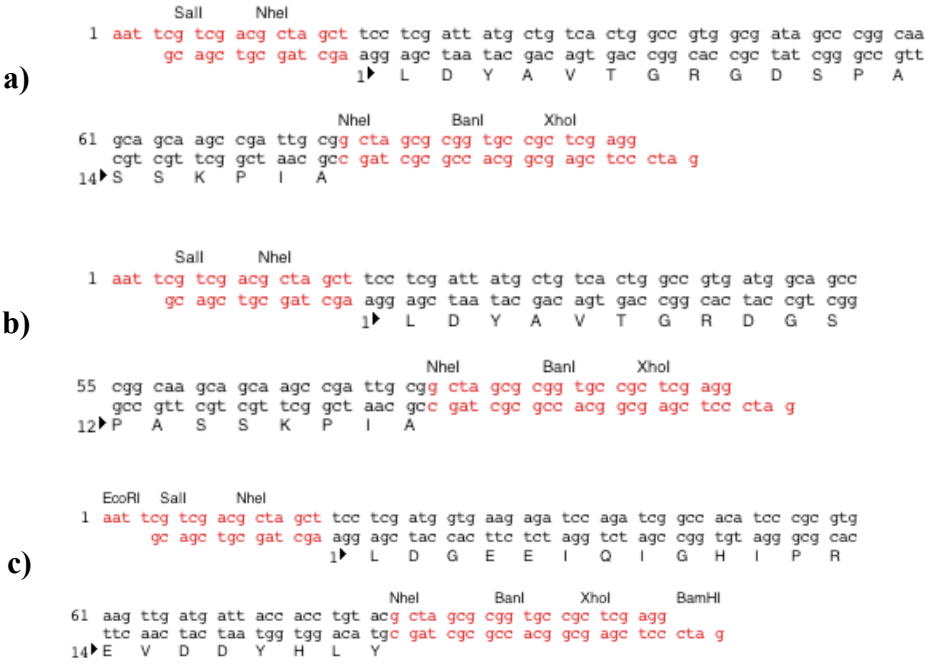


Figure V-3. DNA oligonucleotide sequences encoding **a)** the RGD cell-binding domain and **b)** the scrambled RDG and **c)** scrambled REVD negative control domains were designed and ordered commercially.

digestion and ligated into *Ban* I-linearized pEC2 vector [6] to produce pEC2-SC5, pEC2-RGD, and pEC2-RDG.

ELF monomer (Figure V-4) was obtained through *Ban* I digestion of pUC19-ELF [1]. The 75-bp fragment was multimerized through ligation with T4 DNA ligase, and the multimerization mixture was run on a 2% agarose gel. The band corresponding to pentamer (375 bp) was cut out and extracted from the gel. The pentamer DNA was ligated into pEC2-SC5 vector that had been digested with *Ban* I and dephosphorylated with CIP to yield pEC2-SC5-ELF_5.

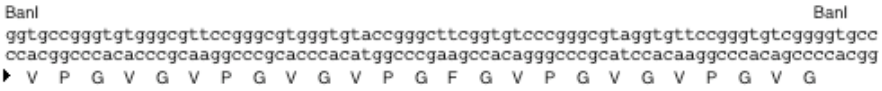


Figure V-4. DNA and protein sequence of the ELF monomer digested from pUC19-ELF.

Meanwhile, the gene encoding the A294G mutant PheRS (phes*) was cut from pUC19-FS [1] with *Sph* I. It was ligated into the *Sph* I site of pET28cyl [5] to create pET28cyl-phes*. The pEC2-SC5-ELF5 vector was digested with *Xho* I / *Spa* I, but attempts to ligate it into the pET28cyl-phes* failed. Sequencing revealed a second *Xho* I site on the pET28cyl-phes* between the *Sph* I ligation site and the start of the phes* gene. Future work will focus on removing this cut site through site-directed mutagenesis.

V.3 Methods

*V.3.1 pET28cyl-phes**

The phes* gene was isolated from pUC19-FS [1] through digestion with *Sph* I (NEB). It was ligated using T4 DNA ligase (NEB, room temperature, overnight) into pET28cyl [5] that had been linearized with *Sph* I (37°C, overnight) to produce pET28cyl-phes*. The A294G mutation was confirmed through DNA sequencing.

V.3.2 pEC2-RGD, RDG, and SC5

The cell-binding domain oligonucleotides (Figure V-4) were ordered from Qiagen and resuspended at 5 μg/μL in 10 mM Tris buffer at pH 8.1. A 10-μL aliquot of each strand was added to 80 μL of annealing buffer (100 mM NaCl, 20 mM $MgCl_2$) and

heated to near boiling for 2 to 3 minutes. The solution was allowed to cool slowly back to room temperature and run on a 2 % agarose gel. The bands corresponding to the expected DNA (SC5 = 105 bp, RGD and RDG = 99 bp) were cut from the gel and extracted using a QIAquick Gel Extraction Kit (Qiagen).

The cell-binding domain fragments were phosphorylated using T4 polynucleotide kinase (NEB). Purified DNA was incubated at 37°C for 2 h with kinase in the supplied kinase buffer. The phosphorylated cell-binding domains were ligated (T4 DNA ligase [NEB], room temperature, 2 h) into pUC19 vector [1] that had been digested at 37°C overnight with *Eco*R I (NEB) and *Bam*H I (NEB). The pUC19-RGD, RDG, and SC5 vectors were transformed into XL-1 Blue competent cells (Stratagene).

The XL-1 Blue strains harboring the plasmids pUC19-RGD, RDG, and SC5 were grown in 5 mL overnight 2xYT cultures, and the plasmid DNA was isolated using a Miniprep Kit (Qiagen). The vectors were digested at 37°C overnight with *Eco*R I and *Bam*H I, the digestion mixtures were run on a 2 % agarose gel, and the appropriate cell-binding domain fragments were excised and extracted using a QIAquick Kit. Purified fragments were ligated into pEC2 plasmid [7] that had been digested with *Eco*R I and *Bam*H I to create pEC2-RGD, RDG, and SC5.

V.3.3 pEC2-SC5-ELF$_5$ and ELF$_6$

The DNA fragment encoding one repeat of the desired elastin-like region with Phe sites, designated ELF (Figure V-5), was obtained through digestion of pUC19-ELF (Nandita Sharma). The plasmid DNA was isolated from a 500-mL overnight 2xYT

culture using a MaxiPrep Kit (Qiagen) and digested at 37°C overnight with *Ban* I (NEB). The digestion mixture was run on a 3 % agarose gel, the band corresponding to ELF monomer (75 bp) was excised, and the DNA was extracted with a Zymoclean Gel DNA Recovery Kit (Zymo Research). Two separate strategies were employed to multimerize the ELF monomer and insert it into the pEC2-SC5 plasmid. In the first scheme, the ELF monomer was self-ligated with T4 DNA ligase in the supplied ligase buffer at 0°C for <5 min before the ligation mixture was immediately loaded onto a 2 % agarose gel. The 375-bp band, corresponding to and ELF pentamer (ELF_5), was excised from the gel, and the DNA was extracted with a QIAquick Kit. The pEC2-SC5 plasmid was digested with *Ban* I (37°C, overnight) and dephosphorylated with calf intestine phosphatase (CIP, NEB, room temperature, 5 min) before being purified on a 2 % agarose gel. Linearized and dephosphorylated pEC2-SC5 was ligated for 3 h at room temperature with the purified ELF pentamer fragment, and the ligation mixtures were transformed into JM109 competent cells (Stratagene). Transformation colonies were test-digested with *Xho* I and *Sal* I; a band is expected at 468 bp if ELF pentamer successfully inserted into pEC2-SC5. A colony was isolated that carried the desired plasmid, designated pEC2-SC5-ELF_5.

In the second procedure attempted to produce pEC2-SC5-ELF_5, the linearized and dephosphorylated pEC2-SC5 plasmid was added directly to the ELF monomer ligation mixture 30 s after ligation was initiated by the addition of T4 DNA ligase. The ligation was allowed to proceed for 3 h before transformation of the mixture into JM109

competent cells. A colony was isolated that harbored a plasmid designated pEC2-SC5-ELF$_6$; the multimer that had been inserted was the hexamer of ELF.

V.4 Results

Progress was made in the attempt to clone an artificial extracellular matrix protein containing the RGD cell-binding domain from fibronectin and regular Phe residues for incorporation of the analogue *p*-azidophenylalanine (RGD-F) along with negative control proteins RDG-F and SC5-F. The cloning strategy is depicted in Figure V-2.

A pET28cyl-phes* vector containing the A294G mutant phenylalanyl-tRNA synthetase was created and confirmed through sequencing. ELF monomer, isolated from the digestion of pUC19-ELF, was successfully multimerized with T4 ligase (Figure V-5).

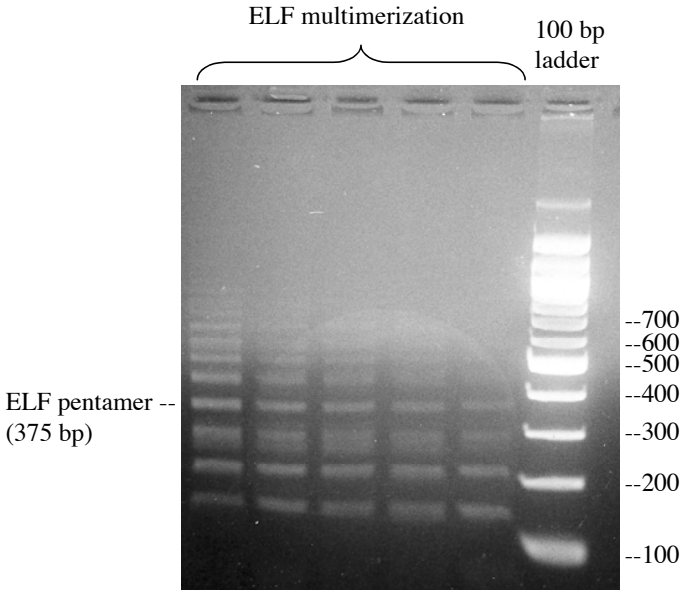


Figure V-5. Representative 3% agaraose gel of a multimerization reaction. ELF monomer DNA (with *Ban* I sticky ends) was incubated with T4 DNA ligase at 0°C for 3 min before being loaded onto the gel.

Excising the 375-bp band from the agarose gel and ligating the extracted DNA into a linearized pEC2-SC5 cloning vector resulted in a colony harboring a pEC2-SC5-ELF_5 (pentamer, Figure V-6). In a separate experiment, ELF monomer was multimerized in the presence of linearized pEC2-SC5 vector, followed by transformation of the crude multimerization mixture. This method produced a colony harboring pEC2-SC5-ELF_6 (hexamer, Figure V-6).

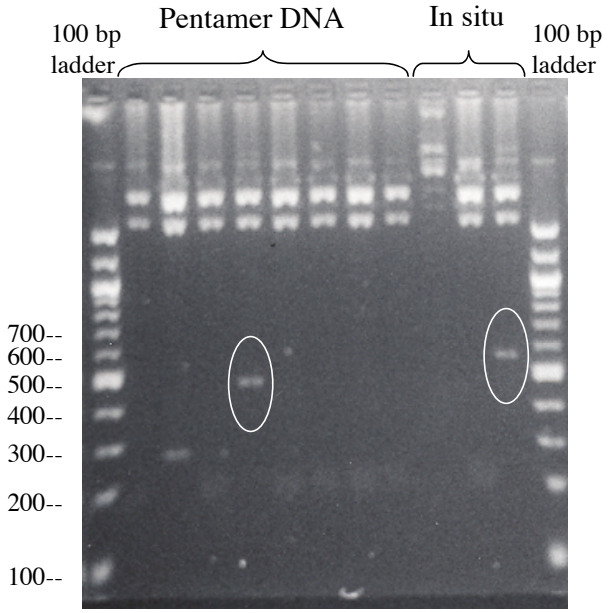


Figure V-6. Test digestions (with *Xho* I and *Sal* I) of colonies of cells transformed with ligation mixtures of pEC2-SC5/*Ban* I and ELF monomer with *Ban* I sticky ends. Transformants in lanes 2 to 9 resulted from ligations of pEC2-SC5/*Ban* I with ELF pentamer DNA that had been excised from a 3% agarose gel of the ELF multimerization. The circled band in lane 5 indicates presence of pEC2-SC5-ELF_5 (pentamer). Transformants in lanes 10 to 12 resulted from *in situ* ligations of pEC2-SC5/*Ban* I with ELF monomer. The circled band in lane 12 indicates the presence of pEC2-SC5-ELF_6 (hexamer).

V.5 Conclusions

The progress made in cloning the desired artificial extracellular matrix protein constructs, especially in the difficult multimerization step, will facilitate future production of these proteins. The proteins, one containing the RGD cell-binding domain, and negative control proteins for both the RGD and CS5 cell-binding domains, will enable studies of cellular behavior on photocrosslinked substrates through the incorporation of the photoreactive amino acid *p*-azidophenylalanine.

V.6 References

1. Sharma N: **Biosynthetic introduction of aryl bromide functionality into proteins [Doctoral thesis]**. Amherst: University of Massachusetts: 2001.

2. Hubbell JA, Massia SP, Desai NP, Drumheller PD: **Endothelial Cell-Selective Materials for Tissue Engineering in the Vascular Graft Via a New Receptor**. *Bio-Technology* 1991, **9**:568-572.

3. Liu JC, Heilshorn SC, Tirrell DA: **Comparative cell response to artificial extracellular matrix proteins containing the RGD and CS5 cell-binding domains**. *Biomacromolecules* 2004, **5**:497-504.

4. Pierschbacher MD, Ruoslahti E: **Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule**. *Nature* 1984, **309**:30-33.

5. Liu CY, Apuzzo MLJ, Tirrell DA: **Engineering of the extracellular matrix: Working toward neural stem cell programming and neurorestoration - Concept and progress report**. *Neurosurgery* 2003, **52**:1154-1165.

6. Cantor EJ: **[Doctoral thesis]**. Amherst, MA: University of Massachusetts: 1997.

7. Panitch A: **Design, synthesis and characterization of artificial extracellular matrix proteins for tissue engineering [Doctoral thesis]**. Amherst, MA: University of Massachusetts at Amherst: 1997.

Appendix A

DNA SEQUENCES AND PLASMID MAPS

TABLE OF CONTENTS

A.1 pETcyl-phes* plasmid information

Submitted by: Stacey Maskarinec, Marissa Mock **Date 4/22/04**

Strain name: **XL-1 blue (***SupE*44 *hsdR*17 *recA*1 *endA*1 *gyrA*46 *thi relA*1 *lac*$^-$ F'[*proAB*$^+$ *lacI*^{q} l*acZ*ΔM15 Tn*10*(*tet*^{r})] (From Stratagene) **) / pET28cyl-phes***

Vector (kb): **pET28cyl-phes* (7.0)**

Cloning site: See plasmid map.

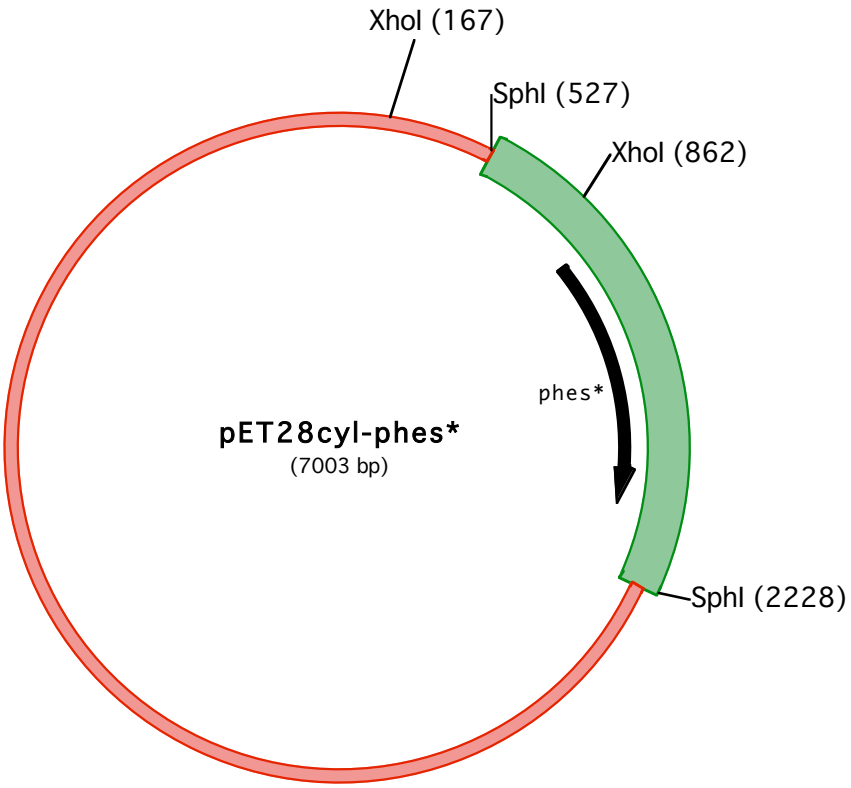
Construction of pET28cyl-phes* plasmid:

The phes***** gene was cut from pUC19-FS (Nandita Sharma) with Sph I. It was ligated into the Sph I site of pET28cyl (Charles Liu).

Source available :

12 % Glycerol culture in Marissa-3 freezer box
Culture conditions: 2xYT, 37°C

A.1.1 Plasmid map



A.1.2 Position of elements

	bp
Vector size	7003
Full insert region	532-2233
PheRS* gene	992-1073

A.1.3 Full sequence (pET28cyl-phes)*

```
  1 ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATT
    TAGGCCTATATCAAGGAGGAAAGTCGTTTTTTGGGGAGTTCTGGGCAAATCTCCGGGGTTCCCCAATACGATCAATAA

 79 GCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGT
    CGAGTCGCCACCGTCGTCGGTTGAGTCGAAGGAAAGCCCGAAACAATCGTCGGCCTAGAGTCACCACCACCACCACCA

                Xhol (167)
157 GCTCGA ctt act cga gaa gct t CAT GTG GTG GTG GTG GTG GTG ACC CAT TTG CTG TCC
    CGAGCT gaa tga gct ctt cga a GTA CAC CAC CAC CAC CAC CAC TGG GTA AAC GAC AGG

215 ACC AGT CAT GCT CGC CAT gGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTAT
    TGG TCA GTA CGA GCG GTA cCATATAGAGGAAGAATTTCAATTTGTTTTAATAAAGATCTCCCCTTAACAATA

287 CCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCGATCCTCTACGCCGGACGCATCGT
    GGCGAGTGTTAAGGGGATATCACTCAGCATAATTAAAGCGCCCTAGCTCTAGAGCTAGGAGATGCGGCCTGCGTAGCA

365 GGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCG
    CCGGCCGTAGTGGCCGCGGTGTCCACGCCAACGACCGCGGATATAGCGGCTGTAGTGGCTACCCCTTCTAGCCCGAGC

443 CCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCAT
    GGTGAAGCCCGAGTACTCGCGAACAAAGCCGCACCCATACCACCGTCCGGGGCACCGGCCCCCTGACAACCCGCGGTA

          SphI (527)
521 CTCCTTGCATGccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcct
    GAGGAACGTACggcttgctggctcgcgtcgctcagtcactcgctccttcgccttctcgcgggttatgcgtttggcgga

599 ctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaac
    gaggggcgcgcaaccggctaagtaattacgtcgaccgtgctgtccaaagggctgacctttcgcccgtcactcgcgttg

677 gcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtgga
    cgttaattacactcaatcgagtgagtaatccgtggggtccgaaatgtgaaatacgaaggccgagcatacaacacacct

755 attgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagctcgaaattaaccctcactaa
    taacactcgcctattgttaaagtgtgtcctttgtcgatactggtactaatgcggttcgagctttaattgggagtgatt

                                Xhol (862)
833 agggaacaaaagctggtaccgggccccccctcgaggtcgacggtatcgataagcttgatatcgaattcccccgggacc
    tcccttgttttcgaccatggcccgggggggagctccagctgccatagctattcgaactatagcttaaggggccctgg

911 aaaatggcaagtaaaatagcctgatgggataggctctaagtccaacgaaccagtgtcaccactg
    ttttaccgttcattttatcggactaccctatccgagattcaggttgcttggtcacagtggtgac
```

```
 975 acacaatgaggaaaaccatgtcacatctcgcagaactggttgccagtgcgaaggcggccattagccaggcgtcagat
     tgtgttactccttttggtacagtgtagagcgtcttgaccaacggtcacgcttccgccggtaatcggtccgcagtcta
                   1▶ M  S  H  L  A  E  L  V  A  S  A  K  A  A  I  S  Q  A  S  D

1052 gttgccgcgttagataatgtgcgcgtcgaatatttgggtaaaaaagggcacttaacccttcagatgacgaccctgcg
     caacggcgcaatctattacacgcgcagcttataaacccatttttcccgtgaattgggaagtctactgctgggacgc
  21▶ V  A  A  L  D  N  V  R  V  E  Y  L  G  K  K  G  H  L  T  L  Q  M  T  T  L  R

1129 tgagctgccgccagaagagcgtccggcagctggtgcggttatcaacgaagcgaaagagcaggttcagcaggcgctga
     actcgacggcggtcttctcgcaggccgtcgaccacgccaatagttgcttcgctttctcgtccaagtcgtccgcgact
  46▶  E  L  P  P  E  E  R  P  A  A  G  A  V  I  N  E  A  K  E  Q  V  Q  Q  A  L

1206 atgcgcgtaaagcggaactggaaagcgctgcactgaatgcgcgtctggcggcggaaacgattgatgtctctctgcca
     tacgcgcatttcgccttgacctttcgcgacgtgacttacgcgcagaccgccgcctttgctaactacagagagacggt
  72▶N  A  R  K  A  E  L  E  S  A  A  L  N  A  R  L  A  A  E  T  I  D  V  S  L  P

1283 ggtcgtcgcattgaaaacggcggtctgcatccggttacccgtaccatcgaccgtatcgaaagtttcttcggtgagct
     ccagcagcgtaacttttgccgccagacgtaggccaatgggcatggtagctggcatagctttcaaagaagccactcga
  98▶ G  R  R  I  E  N  G  G  L  H  P  V  T  R  T  I  D  R  I  E  S  F  F  G  E  L

1360 tggctttaccgtggcaaccgggccggaaatcgaagacgattatcataacttcgatgctctgaacattcctggtcacc
     accgaaatggcaccgttggcccggcctttagcttctgctaatagtattgaagctacgagacttgtaaggaccagtgg
 123▶  G  F  T  V  A  T  G  P  E  I  E  D  D  Y  H  N  F  D  A  L  N  I  P  G  H

1437 acccggcgcgcgctgaccacgacactttctggtttgacactacccgcctgctgcgtacccagacctctggcgtacag
     tgggccgcgcgcgactggtgctgtgaaagaccaaactgtgatgggcggacgacgcatgggtctggagaccgcatgtc
 149▶H  P  A  R  A  D  H  D  T  F  W  F  D  T  T  R  L  L  R  T  Q  T  S  G  V  Q

1514 atccgcaccatgaaagcccagcagccaccgattcgtatcatcgcgcctggccgtgtttatcgtaacgactacgacca
     taggcgtggtactttcgggtcgtcggtggctaagcatagtagcgcggaccggcacaaatagcattgctgatgctggt
 175▶ I  R  T  M  K  A  Q  Q  P  P  I  R  I  I  A  P  G  R  V  Y  R  N  D  Y  D  Q

1591 gactcacacgccgatgttccatcagatggaaggtctgattgttgataccaacatcagctttaccaacctgaaaggca
     ctgagtgtgcggctacaaggtagtctaccttccagactaacaactatggttgtagtcgaaatggttggactttccgt
 200▶  T  H  T  P  M  F  H  Q  M  E  G  L  I  V  D  T  N  I  S  F  T  N  L  K  G

1668 cgctgcacgacttcctgcgtaacttctttgaggaagatttgcagattcgcttccgtccttcctacttcccgtttacc
     gcgacgtgctgaaggacgcattgaagaaactccttctaaacgtctaagcgaaggcaggaaggatgaagggcaaatgg
 226▶T  L  H  D  F  L  R  N  F  F  E  E  D  L  Q  I  R  F  R  P  S  Y  F  P  F  T

1745 gaaccttctgcagaagtggacgtcatgggtaaaaacggtaaatggctggaagtgctgggctgcgggatggtgcatcc
     cttggaagacgtcttcacctgcagtacccatttttgccatttaccgaccttcacgacccgacgccctaccacgtagg
 252▶ E  P  S  A  E  V  D  V  M  G  K  N  G  K  W  L  E  V  L  G  C  G  M  V  H  P

1822 gaacgtgttgcgtaacgttggcatcgacccggaagtttactctggtttcggcttcgggatggggatggagcgtctga
     cttgcacaacgcattgcaaccgtagctgggccttcaaatgagaccaaagccgaagccctacccctacctcgcagact
 277▶  N  V  L  R  N  V  G  I  D  P  E  V  Y  S  G  F  G  F  G  M  G  M  E  R  L

1899 ctatgttgcgttacggcgtcaccgacctgcgttcattcttcgaaaacgatctgcgtttcctcaaacagtttaaataa
     gatacaacgcaatgccgcagtggctggacgcaagtaagaagcttttgctagacgcaaaggagtttgtcaaatttatt
 303▶T  M  L  R  Y  G  V  T  D  L  R  S  F  F  E  N  D  L  R  F  L  K  Q  F  K

1976 ggcaggaatagattatgaaattcagtgaactgtggttacgcgaatgggtgaacccggcgattgatagcgatgcgctg
     ccgtccttatctaatactttaagtcacttgacaccaatgcgcttacccacttgggccgctaactatcgctacgcgac
```

```
1976 ggcaggaatagattatgaaattcagtgaactgtggttacgcgaatgggtgaacccggcgattgatagcgatgcgctg
     ccgtccttatctaatactttaagtcacttgacaccaatgcgcttacccacttgggccgctaactatcgctacgcgac

2053 gcaaatcaaatcactatggcgggcctggaagttgggggatccactagttctagagcggccgccaccgcggtggagct
     cgtttagtttagtgataccgcccggaccttcaacccccctaggtgatcaagatctcgccggcggtggcgccacctcga

2130 ccaattcgccctatagtgagtcgtattacaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggc
     ggttaagcgggatatcactcagcataatgttaagtgaccggcagcaaaatgttgcagcactgacccttttgggaccg

                          SphI (2228)
2207 gttacccaacttaatcgccttgcatgCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCT
     caatgggttgaattagcggaacgtacGTGGTAAGGAACGCCGCCGCCACGAGTTGCCGGAGTTGGATGATGACCCGA

2284 GCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGAGATCCCGGACACCATCGAATGGCGCAAAACCTTTCGCGG
     CGAAGGATTACGTCCTCAGCGTATTCCCTCTCGCAGCTCTAGGGCCTGTGGTAGCTTACCGCGTTTTGGAAAGCGCC

2361 TATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAG
     ATACCGTACTATCGCGGGCCTTCTCTCAGTTAAGTCCCACCACTTACACTTTGGTCATTGCAATATGCTACAGCGTC

2438 AGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAA
     TCATACGGCCACAGAGAATAGTCTGGCAAAGGGCGCACCACTTGGTCCGGTCGGTGCAAAGACGCTTTTGCGCCCTT

2515 AAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTT
     TTTCACCTTCGCCGCTACCGCCTCGACTTAATGTAAGGGTTGGCGCACCGTGTTGTTGACCGCCCGTTTGTCAGCAA

2592 GCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCG
     CGACTAACCGCAACGGTGGAGGTCAGACCGGGACGTGCGCGGCAGCGTTTAACAGCGCCGCTAATTTAGAGCGCGGC

2669 ATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAAT
     TAGTTGACCCACGGTCGCACCACCACAGCTACCATCTTGCTTCGCCGCAGCTTCGGACATTTCGCCGCCACGTGTTA

2746 CTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGC
     GAAGAGCGCGTTGCGCAGTCACCCGACTAGTAATTGATAGGCGACCTACTGGTCCTACGGTAACGACACCTTCGACG

2823 CTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAG
     GACGTGATTACAAGGCCGCAATAAAGAACTACAGAGACTGGTCTGTGGGTAGTTGTCATAATAAAAGAGGGTACTTC

2900 ACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGT
     TGCCATGCGCTGACCCGCACCTCGTAGACCAGCGTAACCCAGTGGTCGTTTAGCGCGACAATCGCCCGGGTAATTCA

2977 TCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACG
     AGACAGAGCCGCGCAGACGCAGACCGACCGACCGTATTTATAGAGTGAGCGTTAGTTTAAGTCGGCTATCGCCTTGC

3054 GGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGA
     CCTTCCGCTGACCTCACGGTACAGGCCAAAAGTTGTTTGGTACGTTTACGACTTACTCCCGTAGCAAGGGTGACGCT

3131 TGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGAT
     ACGACCAACGGTTGCTAGTCTACCGCGACCCGCGTTACGCGCGGTAATGGCTCAGGCCCGACGCGCAACCACGCCTA

3208 ATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTT
     TAGAGCCATCACCCTATGCTGCTATGGCTTCTGTCGAGTACAATATAGGGCGGCAATTGGTGGTAGTTTGTCCTAAA

3285 TCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGT
     AGCGGACGACCCCGTTTGGTCGCACCTGGCGAACGACGTTGAGAGAGTCCCGGTCCGCCACTTCCCGTTAGTCGACA

3362 TGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGAT
     ACGGGCAGAGTGACCACTTTTCTTTTTGGTGGGACCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTA

3439 TCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGC
     AGTAATTACGTCGACCGTGCTGTCCAAAGGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACATTCAATCG

3516 TCACTCATTAGGCACCGGGATCTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGG
     AGTGAGTAATCCGTGGCCCTAGAGCTGGCTACGGGAACTCTCGGAAGTTGGGTCAGTCGAGGAAGGCCACCCGCGCC

3593 GGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTG
     CCGTACTGATAGCAGCGGCGTGAATACTGACAGAAGAAATAGTACGTTGAGCATCCTGTCCACGGCCGTCGCGAGAC

3670 GGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGC
     CCAGTAAAAGCCGCTCCTGGCGAAAGCGACCTCGCGCTGCTACTAGCCGGACAGCGAACGCCATAAGCCTTAGAACG

3747 ACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATG
     TGCGGGAGCGAGTTCGGAAGCAGTGACCAGGGCGGTGGTTTGCAAAGCCGCTCTTCGTCCGGTAATAGCGGCCGTAC

3824 GCGGCCCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTA
     CGCCGGGGTGCCCACGCGTACTAGCACGAGGACAGCAACTCCTGGGCCGATCCGACCGCCCCAACGGAATGACCAAT

3901 GCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACAT
     CGTCTTACTTAGTGGCTATGCGCTCGCTTGCACTTCGCTGACGACGACGTTTTGCAGACGCTGGACTCGTTGTTGTA

3978 GAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATC
     CTTACCAGAAGCCAAAGGCACAAAGCATTTCAGACCTTTGCGCCTTCAGTCGCGGGACGTGGTAATACAAGGCCTAG
```

```
4055 TGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGT
     ACGTAGCGTCCTACGACGACCGATGGGACACCTTGTGGATGTAGACATAATTGCTTCGCGACCGTAACTGGGACTCA

4132 GATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATGTTCAT
     CTAAAAAGAGACCAGGGCGGCGTAGGTATGGCGGTCAACAAATGGGAGTGTTGCAAGGTCATTGGCCCGTACAAGTA

4209 CATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTTAC
     GTAGTCATTGGGCATAGCACTCGTAGGAGAGAGCAAAGTAGCCATAGTAATGGGGGTACTTGTCTTTAGGGGGAATG

4286 ACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCT
     TGCCTCCGTAGTCACTGGTTTGTCCTTTTTTGGCGGGAATTGTACCGGGCGAAATAGTCTTCGGTCTGTAATTGCGA

4363 TCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGC
     AGACCTCTTTGAGTTGCTCGACCTGCGCCTACTTGTCCGTCTGTAGACACTTAGCGAAGTGCTGGTGCGACTACTCG

4440 TTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACA
     AAATGGCGTCGACGGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGT

4517 GCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGC
     CGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGCG

4594 AGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAG
     TCGGTACTGGGTCAGTGCATCGCTATCGCCTCACATATGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTC

4671 AGTGCACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTC
     TCACGTGGTATATACGCCACACTTTATGGCGTGTCTACGCATTCCTCTTTTATGGCGTAGTCCGCGAGAAGGCGAAG

4748 CTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGT
     GAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCA

4825 TATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
     ATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTACACTCGTTTTCCGGTCGTTTTCCGGTCCTTGGCATTTTTC

4902 GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTG
     CGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCAC

4979 GCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC
     CGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGG

5056 TGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTAT
     ACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTATCGAGTGCGACATCCATA

5133 CTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
     GAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAA

5210 ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGA
     TAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCT

5287 TTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACA
     AATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGT

5364 GTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAAC
     CATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTG

5441 CACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
     GTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAA

5518 TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAACAATAAAACT
     ACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTTGTTATTTTGA

5595 GTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGAT
     CAGACGAATGTATTTGTCATTATGTTCCCCACAATACTCGGTATAAGTTGCCCTTTGCAGAACGAGATCCGGCGCTA

5672 TAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATC
     ATTTAAGGTTGTACCTACGACTAAATATACCCATATTTACCCGAGCGCTATTACAGCCCGTTAGTCCACGCTGTTAG

5749 TATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTAC
     ATAGCTAACATACCCTTCGGGCTACGCGGTCTCAACAAAGACTTTGTACCGTTTCCATCGCAACGGTTACTACAATG

5826 AGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTG
     TCTACTCTACCAGTCTGATTTGACCGACTGCCTTAAATACGGAGAAGGCTGGTAGTTCGTAAAATAGGCATGAGGAC

5903 ATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGT
     TACTACGTACCAATGAGTGGTGACGCTAGGGGCCCTTTTGTCGTAAGGTCCATAATCTTCTTATAGGACTAAGTCCA

5980 GAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAG
     CTTTTATAACAACTACGCGACCGTCACAAGGACGCGGCCAACGTAAGCTAAGGACAAACATTAACAGGAAAATTGTC

6057 CGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACG
     GCTAGCGCATAAAGCAGAGCGAGTCCGCGTTAGTGCTTACTTATTGCCAAACCAACTACGCTCACTAAAACTACTGC
```

```
6134 AGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTC
     TCGCATTACCGACCGGACAACTTGTTCAGACCTTTCTTTACGTATTTGAAAACGGTAAGAGTGGCCTAAGTCAGCAG

6211 ACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGT
     TGAGTACCACTAAAGAGTGAACTATTGGAATAAAAACTGCTCCCCTTTAATTATCCAACATAACTACAACCTGCTCA

6288 CGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAAC
     GCCTTAGCGTCTGGCTATGGTCCTAGAACGGTAGGATACCTTGACGGAGCCACTCAAAAGAGGAAGTAATGTCTTTG

6365 GGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTC
     CCGAAAAAGTTTTTATACCATAACTATTAGGACTATACTTATTTAACGTCAAAGTAAACTACGAGCTACTCAAAAAG

6442 TAAGAATTAATTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTC
     ATTCTTAATTAAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAG

6519 CCCGAAAAGTGCCACCTGAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCA
     GGGCTTTTCACGGTGGACTTTAACATTTGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGT

6596 TTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT
     AAAAAATTGGTTATCCGGCTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACA

6673 TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG
     AGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGC

6750 ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCT
     TACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGGGA

6827 AAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGG
     TTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC

6904 AGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGC
     TCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGGCGGCGCGAATTACGCGGCG

6981 TACAGGGCGCGTCCCATTCGCCA
     ATGTCCCGCGCAGGGTAAGCGGT
```

A.2 pEC2-SC5, -RGD, -RDG plasmid information

Submitted by Marissa Mock **Date 4/22/04**

Strain name/ **XL-1 blue (***SupE*44 *hsdR*17 *recA*1 *endA*1 *gyrA*46 *thi relA*1 *lac*$^-$ F'[*proAB*$^+$ *lacI*^{q} l*acZ*ΔM15 Tn*10*(*tet*^{r})] (From Stratagene) **)/ pEC2-SC5**
pEC2-RGD
pEC2-RDG

Vector (kb) : **pEC2-SC5**
pEC2-RGD
pEC2-RDG

Cloning site: See plasmid map.

Construction of pEC2-CBD plasmids

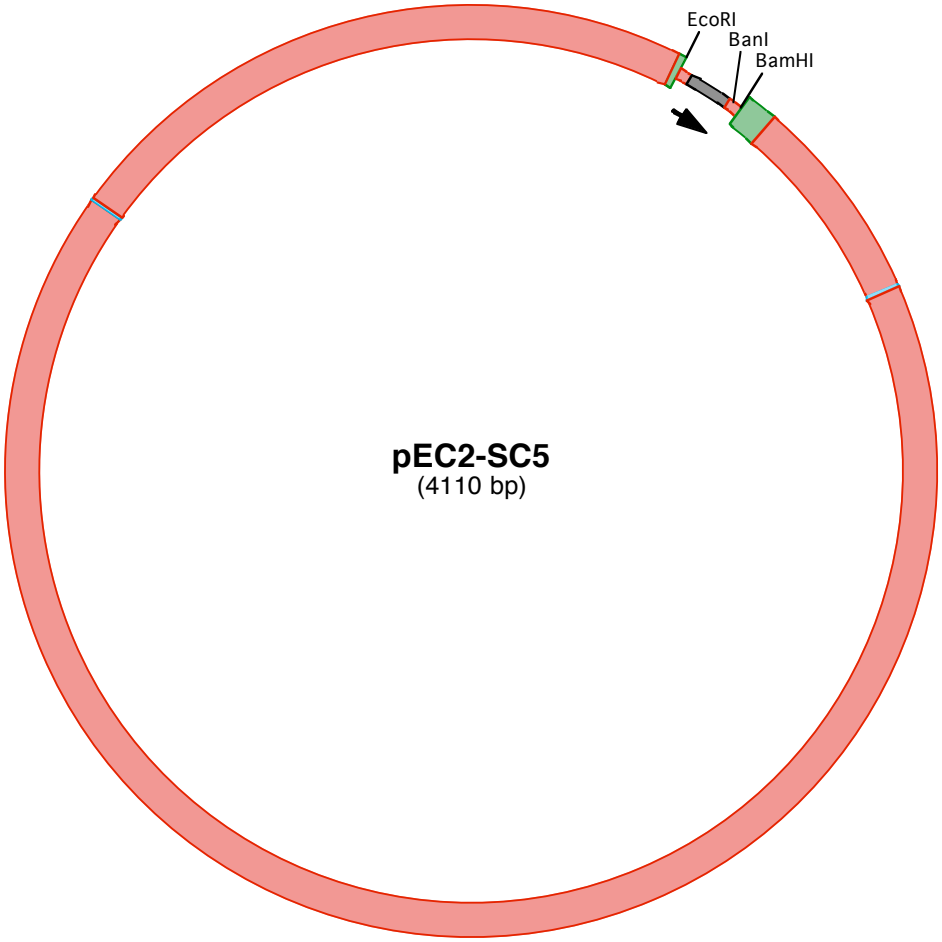
The cell binding domain region was cut out of the appropriate pUC19-SC5, pUC19-RGD, or pUC19-RDG vector with EcoR I/ BamH I digestion. The pEC2 (Eric Cantor) vector was cut with EcoR I/BamH I and ligated with the inserts to produce pEC2-SC5, pEC2-RGD, and pEC2-RDG.

Plasmid map and full sequence are shown only for pEC2-SC5; others are similar.

Source available :

12 % Glycerol culture in Marissa freezer box
Culture conditions: 2xYT, 37°C

A.2.1 Plasmid map



A.2.2 Position of elements

	bp
Vector size	4110
SC5 insert	314-420
RGD insert	314-414
RDG insert	314-414

A.2.3 Full sequence (pEC2-SC5)

```
   1 GTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAA
     CAACTGCGGCCCGTTCTCGTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTT

  83 AGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACT
     TCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACGCCGGTTGAATGA

 165 TCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGG
     AGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATTGAGCGGAACTAGCAACC

                                                                 PstI            EcoRI
 247 GAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAgatctgatcag aat tcg tcg
     CTTGGCCTCGACTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACGTctagactagtc tta agc agc

       NheI
 324 acg cta gct tcc tcg atg gtg aag aga tcc aga tcg gcc aca tcc cgc gtg aag ttg atg
     tgc gat cga agg agc tac cac ttc tct agg tct agc cgg tgt agg gcg cac ttc aac tac
               1▶  L   D   G   E   E   I   Q   I   G   H   I   P   R   E   V   D

                         NheI         BanI       XhoI      BamHI
 384 att acc acc tgt acg cta gcg cgg tgc cgc tcg agg gatccatctagagtcgacgtcggccgttaaccta
     taa tgg tgg aca tgc gat cgc gcc acg gcg agc tcc ctaggtagatctcagctgcagccggcaattggat
  17▶D   Y   H   L   Y

           PstI
 454 ggagatctgcaGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAAT
     cctctagacgtCGTTACCGTTGTTGCAACGCGTTTGATAATTGACCGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTA

 536 AGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCT
     TCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGAGCCGGGAAGGCCGACCGACCAAATAACGACTATTTAGA

 618 GGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
     CCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGGGCATAGCATCAATAGATGT

 700 CGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCgtgcCTCACTGATTAAGCATTGGT
     GCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGcacgGAGTGACTAATTCGTAACCA

 782 AACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGAT
     TTGACAGTCTGGTTCAAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTA

 864 CCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCTTAATAAGATGATCT
     GGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGAATTATTCTACTAGA

 946 TCTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTC
     AGAACTCTAGCAAAACCAGACGCGCATTAGAGAACGAGACTTTTGCTTTTTTGGCGGAACGTCCCGCCAAAAAGCTTCCAAG

1028 TCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTA
     AGACTCGATGGTTGAGAAACTTGGCTCCATTGACCGAACCTCCTCGCGTCAGTGGTTTTGAACAGGAAAGTCAAATCGGAAT

1110 ACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGG
     TGGCCGCGTACTGAAGTTCTGATTGAGGAGATTTAGTTAATGGTCACCGACGACGGTCACCACGAAAACGTACAGAAAGGCC

1192 GTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAG
     CAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCTGACTTGCCCCCAAGCACGTATGTCAGGTCGAACCTC

1274 CGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGA
     GCTTGACGGATGGGCCTTGACTCACAGTCCGCACCTTACTCTGTTTGCGCCGGTATTGTCGCCTTACTGTGGCCATTTGGCT
```

```
1356 AAGGCAGGAACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACC
     TTCCGTCCTTGTCCTCTCGCGTGCTCCCTCGGCGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGG

1438 ACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCACT
     TGACTAAACTCGCAGTCTAAAGCACTACGAACAGTCCCCCCGCCTCGGATACCTTTTTGCCGAAACGGCGCCGGGAGAGTGA

1520 TCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGA
     AGGGACAATTCATAGAAGGACCGTAGAAGGTCCTTTAGAGGCGGGGCAAGCATTCGGTAAAGGCGAGCGGCGTCAGCTTGCT

1602 CCGAGCGTAGCGAGTCAGTGAGCGAGGAAGCGGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTTT
     GGCTCGCATCGCTCAGTCACTCGCTCCTTCGCCTTATATAGGACATAGTGTATAAGACGACTGCGTGGCCACGTCGGAAAAA

1684 TCTCCTGCCACATGAAGCACTTCACTGACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGCTGAGG
     AGAGGACGGTGTACTTCGTGAAGTGACTGTGGGAGTAGTCACGGTTGTATCATTCGGTCATATGTGAGGCGATCGCGACTCC

1766 TCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGG
     AGACGGAGCACTTCTTCCACAACGACTGAGTATGGTCCGGACTTAGCGGGGTAGTAGGTCGGTCTTTCACTCCCTCGGTGCC

1848 TTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTCGGGAA
     AACTACTCTCGAAACAACATCCACCTGGTCAACCACTAAAACTTGAAAACGAAACGGTGCCTTGCCAGACGCAACAGCCCTT

1930 GATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCACGTTGTGTCTCAAAATCTCTGATGTTA
     CTACGCACTAGACTAGGAAGTTGAGTCGTTTTCAAGCTAAATAAGTTGTTTCGGTGCAACACAGAGTTTTAGAGACTACAAT

2012 CATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGC
     GTAACGTGTTCTATTTTTATATAGTAGTACTTGTTATTTTGACAGACGAATGTATTTGTCATTATGTTCCCCACAATACTCG

2094 CATATTCAACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTC
     GTATAAGTTGCCCTTTGCAGAACGAGCTCCGGCGCTAATTTAAGGTTGTACCTACGACTAAATATACCCATATTTACCCGAG

2176 GCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGG
     CGCTATTACAGCCCGTTAGTCCACGCTGTTAGATAGCTAACATACCCTTCGGGCTACGCGGTCTCAACAAAGACTTTGTACC

2258 CAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATC
     GTTTCCATCGCAACGGTTACTACAATGTCTACTCTACCAGTCTGATTTGACCGACTGCCTTAAATACGGAGAAGGCTGGTAG

2340 AAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAG
     TTCGTAAAATAGGCATGAGGACTACTACGTACCAATGAGTGGTGACGCTAGGGGCCCTTTTGTCGTAAGGTCCATAATCTTC

2422 AATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTG
     TTATAGGACTAAGTCCACTTTTATAACAACTACGCGACCGTCACAAGGACGCGGCCAACGTAAGCTAAGGACAAACATTAAC

2504 TCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTT
     AGGAAAATTGTCGCTAGCGCATAAAGCAGAGCGAGTCCGCGTTAGTGCTTACTTATTGCCAAACCAACTACGCTCACTAAAA

2586 GATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCG
     CTACTGCTCGCATTACCGACCGGACAACTTGTTCAGACCTTTCTTTACGTATTCGAAAACGGTAAGAGTGGCCTAAGTCAGC

2668 TCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGG
     AGTGAGTACCACTAAAGAGTGAACTATTGGAATAAAAACTGCTCCCCTTTAATTATCCAACATAACTACAACCTGCTCAGCC

2750 AATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTT
     TTAGCGTCTGGCTATGGTCCTAGAACGGTAGGATACCTTGACGGAGCCACTCAAAAGAGGAAGTAATGTCTTTGCCGAAAAA

2832 CAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGG
     GTTTTTATACCATAACTATTAGGACTATACTTATTTAACGTCAAAGTAAACTACGAGCTACTCAAAAAGATTAGTCTTAACC
```

```
2914 TTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGGCTTTGTTGAATAAATCGAACTTTTGCTGAG
     AATTAACCAACATTGTGACCGTCTCGTAATGCGACTGAACTGCCCTGCCGCCGAAACAACTTATTTAGCTTGAAAACGACTC

2996 TTGAAGGATCAGATCACGCATCTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAATCACCAACTGGTCCA
     AACTTCCTAGTCTAGTGCGTAGAAGGGCTGTTGCGTCTGGCAAGGCACCGTTTCGTTTTCAAGTTTTAGTGGTTGACCAGGT

3078 CCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTCTGGCTGGATGATGGGGCGATTCAGGCCTGGTATGAGTCAGC
     GGATGTTGTTTCGAGAGTAGTTGGCACCGAGGGAGTGAAAGACCGACCTACTACCCCGCTAAGTCCGGACCATACTCAGTCG

3160 AACACCTTCTTCACGAGGCAGACCTCAGCGCTCAAAGATGCAGGGGTAAAAGCTAACCGCATCTTTACCGACAAGGCATCCG
     TTGTGGAAGAAGTGCTCCGTCTGGAGTCGCGAGTTTCTACGTCCCCATTTTCGATTGGCGTAGAAATGGCTGTTCCGTAGGC

3242 GCAGTTCAACAGATCGGGAAGGGCTGGATTTGCTGAGGATGAAGGTGGAGGAAGGTGATGTCATTCTGGTGAAGAAGCTCGA
     CGTCAAGTTGTCTAGCCCTTCCCGACCTAAACGACTCCTACTTCCACCTCCTTCCACTACAGTAAGACCACTTCTTCGAGCT

3324 CCGTCTTGGCCGCGACACCGCCGACATGATCCAACTGATAAAAGAGTTTGATGCTCAGGGTGTAGCGGTTCGGTTTATTGAC
     GGCAGAACCGGCGCTGTGGCGGCTGTACTAGGTTGACTATTTTCTCAAACTACGAGTCCCACATCGCCAAGCCAAATAACTG

3406 GACGGGATCAGTACCGACGGTGATATGGGGCAAATGGTGGTCACCATCCTGTCGGCTGTGGCACAGGCTGAACGCCGGAGGA
     CTGCCCTAGTCATGGCTGCCACTATACCCCGTTTACCACCAGTGGTAGGACAGCCGACACCGTGTCCGACTTGCGGCCTCCT

3488 TCgatcCTAGAGCGCACGAATGAGGGCCGACAGGAAGCAAAGCTGAAAGGAATCAAATTTGGCCGCAGGCGTACCGTGGACA
     AGctagGATCTCGCGTGCTTACTCCCGGCTGTCCTTCGTTTCGACTTTCCTTAGTTTAAACCGGCGTCCGCATGGCACCTGT

3570 GGAACGTCGTGCTGACGCTTCATCAGAAGGGCACTGGTGCAACGGAAATTGCTCATCAGCTCAGTATTGCCCGCTCCACGGT
     CCTTGCAGCACGACTGCGAAGTAGTCTTCCCGTGACCACGTTGCCTTTAACGAGTAGTCGAGTCATAACGGGCGAGGTGCCA

3652 TTATAAAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGA
     AATATTTTAAGAACTTCTGCTTTCCCGGAGCACTATGCGGATAAAAATATCCAATTACAGTACTATTATTACCAAAGAATCT

3734 CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCT
     GCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGA

3816 CATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTT
     GTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCATACTCATAAGTTGTAAAGGCACAGCGGGAA

3898 ATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGT
     TAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTGGGTCTTTGCGACCACTTTCATTTTCTACGACTTCTAGTCA

3980 TGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCC
     ACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAAAAGCGGGGCTTCTTGCAAAAGG

4062 AATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGT
     TTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCA
```

A.3 pEC2-SC5-ELF$_5$ and -ELF$_6$ information

Submitted by Marissa Mock **Date 4/22/04**

Strain name: **JM109** ***(****e14–(McrA–) recA1 endA1 gyrA96 thi-1 hsdR17 (rK– mK+) supE44 relA1 £G(lac-proAB) [F′ traD36 proAB lacIqZ£GM15)*. (From Zymo Research) **)/ pEC2-SC5-ELF$_5$**
or / pEC2-SC5-ELF$_6$

Vector (kb): **pEC2-SC5-ELF$_5$ (X)**
pEC2-SC5-ELF$_6$ (X)

Construction of pEC2-CBD plasmids

ELF monomer (see sequence below) was obtained through Ban I digestion of pUC19-ELF (Nandita Sharma).

pEC2-SC5-ELF$_5$:
The 75 bp fragment was multimerized through ligation with T4 DNA ligase for 3 minutes at 0°C. The ligation mixture was run on a 2% agarose gel, and the band corresponding to pentamer (375 bp) was cut out and extracted from the gel. The pentamer DNA was ligated with pEC2-SC5 vector that had been digested with Ban I and dephosphorylated with CIP, to yield pEC2-SC5-ELF$_5$.

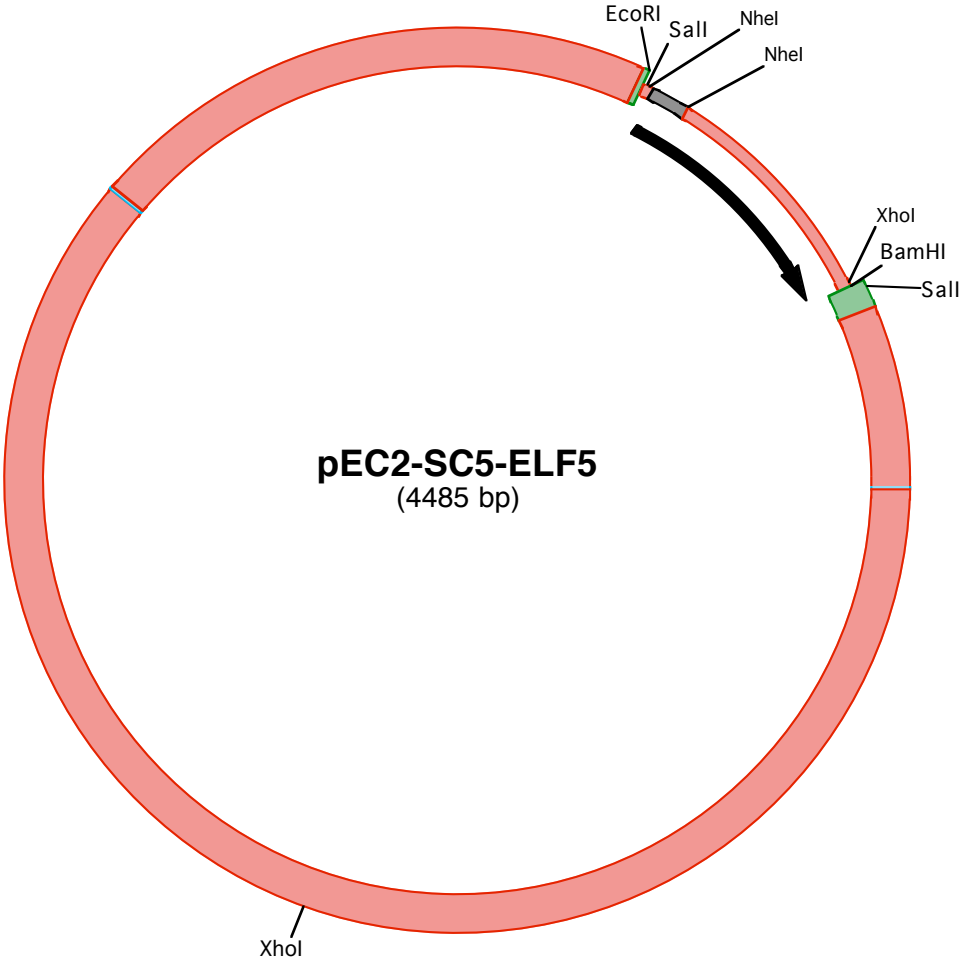
pEC2-SC5-ELF$_6$:
The 75 bp ELF fragment was mixed with T4 ligase at 0°C for 30 s and added directly to pEC2-SC5 vector that had been digested with Ban I and dephosphorylated with CIP. The entire ligation mixture was transformed in JM109 competent cells; screened colonies revealed a strain containing pEC2-SC5-ELF$_6$.

ELF monomer sequence:
ggtgccgggtgtgggcgttccgggcgtgggtgtaccgggcttcggtgtcccgggcgtaggtgttccgggtgtcggggtgcc

Plasmid map and full sequence shown only for pEC2-SC5-ELF$_5$; other is similar.

Source available :
12 % Glycerol culture in Marissa freezer box
Culture conditions: 2xYT, 37°C

A.3.1 Plasmid map



A.3.2 Position of elements

	bp
Vector size	4485
SC5 cell binding domain	314-406
ELF pentamer	407-782

A.3.3 Full sequence (pEC2-SC5-ELF$_5$)

```
   1 GTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACA
     CAACTGCGGCCCGTTCTCGTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGT

  79 GAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCC
     CTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACGCCGG

 157 AACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGC
     TTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATTGAGCG

 235 CTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAgatctgatc
     GAACTAGCAACCCTTGGCCTCGACTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACGTctagactag

          EcoRI            NheI
 313 ag aat tcg tcg acg cta gct tcc tcg atg gtg aag aga tcc aga tcg gcc aca tcc cgc
     tc tta agc agc tgc gat cga agg agc tac cac ttc tct agg tct agc cgg tgt agg gcg
                              1▶ L  D  G  E  E  I  Q  I  G  H  I  P

                                        NheI
 372 gtg aag ttg atg att acc acc tgt acg cta gcg cg gtgccgggtgtgggcgttccgggcgtgggtg
     cac ttc aac tac taa tgg tgg aca tgc gat cgc gc cacggcccacacccgcaaggcccgcacccac
  13▶R  E  V  D  D  Y  H  L  Y  A  S  A  V  P  G  V  G  V  P  G  V  G

 438 taccgggcttcggtgtcccgggcgtaggtgttccgggtgtcggggtgccgggtgtgggcgttccgggcgtgggtgtac
     atggcccgaagccacagggcccgcatccacaaggcccacagccccacggcccacacccgcaaggcccgcacccacatg
  35▶V  P  G  F  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V

 516 cgggcttcggtgtcccgggcgtaggtgttccgggtgtcggggtgccgggtgtgggcgttccgggcgtgggtgtaccgg
     gcccgaagccacagggcccgcatccacaaggcccacagccccacggcccacacccgcaaggcccgcacccacatggcc
  61▶P  G  F  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P

 594 gcttcggtgtcccgggcgtaggtgttccgggtgtcggggtgccgggtgtgggcgttccgggcgtgggtgtaccgggct
     cgaagccacagggcccgcatccacaaggcccacagccccacggcccacacccgcaaggcccgcacccacatggcccga
  87▶G  F  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G

 672 tcggtgtcccgggcgtaggtgttccgggtgtcggggtgccgggtgtgggcgttccgggcgtgggtgtaccgggcttcg
     agccacagggcccgcatccacaaggcccacagccccacggcccacacccgcaaggcccgcacccacatggcccgaagc
 113▶F  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  F

                                                  XhoI      BamHI
 750 gtgtcccgggcgtaggtgttccgggtgtcggg gtg ccg ctc gag g gatccatctagagtcgacgtcggccgt
     cacagggcccgcatccacaaggcccacagccc cac ggc gag ctc c ctaggtagatctcagctgcagccggca
 139▶G  V  P  G  V  G  V  P  G  V  G

 822 taacctaggagatctgcaGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
     attggatcctctagacgtCGTTACCGTTGTTGCAACGCGTTTGATAATTGACCGCTTGATGAATGAGATCGAAGGGCC

 900 CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTT
     GTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGAGCCGGGAAGGCCGACCGACCAAA

 978 ATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
     TAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGG

1056 CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCg
     GCATAGCATCAATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGc
```

```
1134 tgcCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTT
     acgGAGTGACTAATTCGTAACCATTGACAGTCTGGTTCAAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAA

1212 TTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCA
     AATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGT

1290 CTGAGCGTCAGACCCCTTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGA
     GACTCGCAGTCTGGGGAATTATTCTACTAGAAGAACTCTAGCAAAACCAGACGCGCATTAGAGAACGAGACTTTTGCT

1368 AAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGG
     TTTTTGGCGGAACGTCCCGCCAAAAAGCTTCCAAGAGACTCGATGGTTGAGAAACTTGGCTCCATTGACCGAACCTCC

1446 AGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAAT
     TCGCGTCAGTGGTTTTGAACAGGAAAGTCAAATCGGAATTGGCCGCGTACTGAAGTTCTGATTGAGGAGATTTAGTTA

1524 TACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC
     ATGGTCACCGACGACGGTCACCACGAAAACGTACAGAAAGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCG

1602 AGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGC
     TCGCCAGCCTGACTTGCCCCCCAAGCACGTATGTCAGGTCGAACCTCGCTTGACGGATGGGCCTTGACTCACAGTCCG

1680 GTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGG
     CACCTTACTCTGTTTGCGCCGGTATTGTCGCCTTACTGTGGCCATTTGGCTTTCCGTCCTTGTCCTCTCGCGTGCTCC

1758 GAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCG
     CTCGGCGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGTGACTAAACTCGCAGTCTAAAGC

1836 TGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCC
     ACTACGAACAGTCCCCCCGCCTCGGATACCTTTTTGCCGAAACGGCGCCGGGAGAGTGAAGGGACAATTCATAGAAGG

1914 TGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGT
     ACCGTAGAAGGTCCTTTAGAGGCGGGGCAAGCATTCGGTAAAGGCGAGCGGCGTCAGCTTGCTGGCTCGCATCGCTCA

1992 CAGTGAGCGAGGAAGCGGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCCTGCCAC
     GTCACTCGCTCCTTCGCCTTATATAGGACATAGTGTATAAGACGACTGCGTGGCCACGTCGGAAAAAAGAGGACGGTG

2070 ATGAAGCACTTCACTGACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGCTGAGGTCTGCCT
     TACTTCGTGAAGTGACTGTGGGAGTAGTCACGGTTGTATCATTCGGTCATATGTGAGGCGATCGCGACTCCAGACGGA

2148 CGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTG
     GCACTTCTTCCACAACGACTGAGTATGGTCCGGACTTAGCGGGGTAGTAGGTCGGTCTTTCACTCCCTCGGTGCCAAC

2226 ATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTCGGGA
     TACTCTCGAAACAACATCCACCTGGTCAACCACTAAAACTTGAAAACGAAACGGTGCCTTGCCAGACGCAACAGCCCT

2304 AGATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCACGTTGTGTCTCAAAATCTCTGA
     TCTACGCACTAGACTAGGAAGTTGAGTCGTTTTCAAGCTAAATAAGTTGTTTCGGTGCAACACAGAGTTTTAGAGACT

2382 TGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGT
     ACAATGTAACGTGTTCTATTTTTATATAGTAGTACTTGTTATTTTGACAGACGAATGTATTTGTCATTATGTTCCCCA

2460 GTTATGAGCCATATTCAACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGG
     CAATACTCGGTATAAGTTGCCCTTTGCAGAACGAGCTCCGGCGCTAATTTAAGGTTGTACCTACGACTAAATATACCC

2538 TATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAG
     ATATTTACCCGAGCGCTATTACAGCCCGTTAGTCCACGCTGTTAGATAGCTAACATACCCTTCGGGCTACGCGGTCTC

2616 TTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAA
     AACAAAGACTTTGTACCGTTTCCATCGCAACGGTTACTACAATGTCTACTCTACCAGTCTGATTTGACCGACTGCCTT
```

```
2694 TTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGG
     AAATACGGAGAAGGCTGGTAGTTCGTAAAATAGGCATGAGGACTACTACGTACCAATGAGTGGTGACGCTAGGGGCCC

2772 AAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGC
     TTTTGTCGTAAGGTCCATAATCTTCTTATAGGACTAAGTCCACTTTTATAACAACTACGCGACCGTCACAAGGACGCG

2850 CGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGA
     GCCAACGTAAGCTAAGGACAAACATTAACAGGAAAATTGTCGCTAGCGCATAAAGCAGAGCGAGTCCGCGTTAGTGCT

2928 ATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAA
     TACTTATTGCCAAACCAACTACGCTCACTAAAACTACTGCTCGCATTACCGACCGGACAACTTGTTCAGACCTTTCTT

3006 ATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGAC
     TACGTATTCGAAAACGGTAAGAGTGGCCTAAGTCAGCAGTGAGTACCACTAAAGAGTGAACTATTGGAATAAAAACTG

3084 GAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGG
     CTCCCCTTTAATTATCCAACATAACTACAACCTGCTCAGCCTTAGCGTCTGGCTATGGTCCTAGAACGGTAGGATACC

3162 AACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAAT
     TTGACGGAGCCACTCAAAAGAGGAAGTAATGTCTTTGCCGAAAAAGTTTTTATACCATAACTATTAGGACTATACTTA

3240 AAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATT
     TTTAACGTCAAAGTAAACTACGAGCTACTCAAAAAGATTAGTCTTAACCAATTAACCAACATTGTGACCGTCTCGTAA

3318 ACGCTGACTTGACGGGACGGCGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCATCTTC
     TGCGACTGAACTGCCCTGCCGCCGAAACAACTTATTTAGCTTGAAAACGACTCAACTTCCTAGTCTAGTGCGTAGAAG

3396 CCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAATCACCAACTGGTCCACCTACAACAAAGCTCTCATCA
     GGCTGTTGCGTCTGGCAAGGCACCGTTTCGTTTTCAAGTTTTAGTGGTTGACCAGGTGGATGTTGTTTCGAGAGTAGT

3474 ACCGTGGCTCCCTCACTTTCTGGCTGGATGATGGGGCGATTCAGGCCTGGTATGAGTCAGCAACACCTTCTTCACGAG
     TGGCACCGAGGGAGTGAAAGACCGACCTACTACCCCGCTAAGTCCGGACCATACTCAGTCGTTGTGGAAGAAGTGCTC

3552 GCAGACCTCAGCGCTCAAAGATGCAGGGGTAAAAGCTAACCGCATCTTTACCGACAAGGCATCCGGCAGTTCAACAGA
     CGTCTGGAGTCGCGAGTTTCTACGTCCCCATTTTCGATTGGCGTAGAAATGGCTGTTCCGTAGGCCGTCAAGTTGTCT

3630 TCGGGAAGGGCTGGATTTGCTGAGGATGAAGGTGGAGGAAGGTGATGTCATTCTGGTGAAGAAGCTCGACCGTCTTGG
     AGCCCTTCCCGACCTAAACGACTCCTACTTCCACCTCCTTCCACTACAGTAAGACCACTTCTTCGAGCTGGCAGAACC

3708 CCGCGACACCGCCGACATGATCCAACTGATAAAAGAGTTTGATGCTCAGGGTGTAGCGGTTCGGTTTATTGACGACGG
     GGCGCTGTGGCGGCTGTACTAGGTTGACTATTTTCTCAAACTACGAGTCCCACATCGCCAAGCCAAATAACTGCTGCC

3786 GATCAGTACCGACGGTGATATGGGGCAAATGGTGGTCACCATCCTGTCGGCTGTGGCACAGGCTGAACGCCGGAGGAT
     CTAGTCATGGCTGCCACTATACCCCGTTTACCACCAGTGGTAGGACAGCCGACACCGTGTCCGACTTGCGGCCTCCTA

3864 CgatcCTAGAGCGCACGAATGAGGGCCGACAGGAAGCAAAGCTGAAAGGAATCAAATTTGGCCGCAGGCGTACCGTGG
     GctagGATCTCGCGTGCTTACTCCCGGCTGTCCTTCGTTTCGACTTTCCTTAGTTTAAACCGGCGTCCGCATGGCACC

3942 ACAGGAACGTCGTGCTGACGCTTCATCAGAAGGGCACTGGTGCAACGGAAATTGCTCATCAGCTCAGTATTGCCCGCT
     TGTCCTTGCAGCACGACTGCGAAGTAGTCTTCCCGTGACCACGTTGCCTTTAACGAGTAGTCGAGTCATAACGGGCGA

4020 CCACGGTTTATAAAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAAT
     GGTGCCAAATATTTTAAGAACTTCTGCTTTCCCGGAGCACTATGCGGATAAAAATATCCAATTACAGTACTATTATTA
```

```
4098 GGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTC
     CCAAAGAATCTGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAG

4176 AAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCA
     TTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCATACTCATAAGT

4254 ACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAA
     TGTAAAGGCACAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTGGGTCTTTGCGACCACTT

4332 AGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGA
     TCATTTTCTACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACT

4410 GAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGT
     CTCAAAAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCA
```