

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)

© 2005

Marissa Mock

All Rights Reserved

## 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 patterns were produced on a non-adhesive background, endothelial cells selectively adhered to the protein regions to create stable cell patterns.

## TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract .....	iv
Table of Contents .....	vi
List of Tables.....	ix
List of Figures .....	x
Abbreviations .....	xii
Chapter I: Introduction .....	I-1
I.1 Protein biosynthesis.....	I-3
I.1.1 Fidelity of protein biosynthesis .....	I-4
I.2 Protein polymers .....	I-5
I.3 Strategies for the incorporation of noncanonical amino acids.....	I-7
I.3.1 <i>In vitro</i> incorporation.....	I-9
I.3.2 <i>In vivo</i> incorporation.....	I-10
I.3.2.1 Residue-specific incorporation (codon reassignment) .....	I-10
I.3.2.2 Site-specific incorporation.....	I-11
I.3.2.3 Multiple site-specific incorporation .....	I-12
I.4 Thesis objective.....	I-14
I.5 References .....	I-16
Chapter II:	
Stereoselective incorporation of unsaturated isoleucine analogues into proteins <i>in vivo</i>	
.....	II-1
II.1 Abstract .....	II-1
II.2 Introduction.....	II-2
II.3 Methods.....	II-4
II.3.1 Synthesis of analogues .....	II-4
II.3.1.1 General procedures .....	II-4
II.3.1.2 SS- and SR-2-amino-3-methyl-4-pentenoic acid (E-Ile).....	II-5
II.3.1.3 2-amino-3-methyl-4-pentynoic acid (Y-Ile) .....	II-8
II.3.2 Purification of IleRS.....	II-10
II.3.3 ATP/PP <sub>i</sub> exchange.....	II-10
II.3.4 Protein expression.....	II-11
II.3.5 Mass spectrometry .....	II-12
II.3.6 <sup>1</sup> H-NMR spectroscopy .....	II-13
II.4 Results .....	II-13
II.4.1 Synthesis of unsaturated amino acids .....	II-13
II.4.2 Analogue incorporation.....	II-16

II.4.2.1 SS- and SR-Y-Ile .....	II-16
II.4.2.2 SS- and SR-E-Ile.....	II-17
II.4.3 Kinetics of activation of E-Ile by IleRS .....	II-21
II.5 Discussion .....	II-22
II.6 References .....	II-24

### Chapter III:

Controlled degradation of a cell-adhesive, elastomeric protein through incorporation of a fluorinated amino acid .....	III-1
III.1 Abstract .....	III-1
III.2 Introduction .....	III-2
III.3 Methods .....	III-4
III.3.1 5,5,5-trifluoroisoleucine synthesis and purification.....	III-4
III.3.2 Protein expression and purification .....	III-6
III.3.3. Tryptic digest/MALDI.....	III-7
III.3.4 LCST measurement .....	III-7
III.3.5 Analysis of elastase degradation.....	III-7
III.3.6 Cell adhesion .....	III-8
III.4 Results and discussion.....	III-9
III.4.1 Protein synthesis and characterization.....	III-9
III.4.2 Protein degradation .....	III-12
III.4.3 Endothelial cell adhesion .....	III-14
III.5 Supporting information .....	III-17
III.6 References.....	III-18

### Chapter IV:

Lithographic patterning of an intrinsically photoreactive cell-adhesive protein .....	IV-1
IV.1 Abstract .....	IV-1
IV.2 Introduction .....	IV-2
IV.3 Methods .....	IV-3
IV.3.1 Protein expression.....	IV-3
IV.3.2 Protein purification .....	IV-4
IV.3.3 <sup>1</sup> H-NMR of aECM-N <sub>3</sub> .....	IV-5
IV.3.4 FTIR spectroscopy .....	IV-5
IV.3.5 Mechanical testing of bulk films .....	IV-5
IV.1.6 Preparation of aminated glass surfaces .....	IV-6
IV.1.7 Preparation of PEG-modified slides.....	IV-6
IV.1.8 Spin coating of protein films .....	IV-6
IV.1.9 Protein lithography .....	IV-7
IV.1.10 Cell culture.....	IV-7
IV.1.11 Phase contrast and fluorescence microscopy .....	IV-8
IV.1.12 Atomic force microscopy .....	IV-9

IV.4 Results and discussion.....	IV-9-
IV.5 References .....	IV-19
Chapter V:	
Progress toward cloning the gene of an RGD-containing protein	
for photopatterning .....	V-1
V.1 Abstract .....	V-1
V.2 Introduction.....	V-2
V.2.1 Cloning strategy.....	V-3
V.3 Methods.....	V-6
V.3.1 pET28cyl-phen* .....	V-6
V.3.2 pEC2-RGD, RDG, and SC5 .....	V-6
V.3.3 pEC2-SC5-ELF <sub>5</sub> and ELF <sub>6</sub> .....	V-7
V.4 Results .....	V-9
V.5 Conclusions.....	V-11
V.6 References .....	V-11
Appendix: DNA sequences and plasmid maps .....	A-1



## LIST OF TABLES

<i>Number</i>		<i>Page</i>
II-1	Kinetics of activation of E-Ile by IleRS .....	II-22

## LIST OF FIGURES

<i>Number</i>	<i>Page</i>
I-1	The 20 canonical amino acids .....I-2
I-2	The protein biosynthetic machinery .....I-3
I-3	Strategies of noncanonical amino acid incorporation.....I-8
I-4	Noncanonical amino acids incorporated into proteins <i>in vivo</i> .....I-13
II-1	Unsaturated analogs SS-E-Ile, SR-E-Ile, SS-Y-Ile, and SR-Y-Ile ..... II-4
II-2	Synthesis scheme for SS-E-Ile and SR-E-Ile..... II-5
II-3	Mechanism of stereoselectivity in E-Ile preparation ..... II-6
II-4	Synthesis scheme for Y-Ile..... II-9
II-5	<sup>1</sup> H-NMR spectra of SS- and SR-E-Ile after recrystallization..... II-14
II-6	<sup>1</sup> H-NMR spectra of SS- and SR-Y-Ile after recrystallization ..... II-15
II-7	SDS-PAGE and Western blot of proteins incorporating Y-Ile ..... 11-16
II-8	MALDI-TOF spectra of tryptic digest fragments of mDHFR produced in medium supplemented with Y-Ile..... II-17
II-9	SDS-PAGE and Western blot of proteins incorporating SS-E-Ile and SR-E-Ile ..... II-18
II-10	MALDI-TOF spectra of tryptic digest fragments of mDHFR produced in the presence of SS- and SR-E-Ile..... II-19
II-11	<sup>1</sup> H-NMR spectra of proteins incorporating SS- and SR-E-Ile ..... II-20
II-12	SDS-PAGE and Western blot of proteins incorporating varying levels of SS-E-Ile ..... II-21
III-1	Amino acid sequences of aECM proteins ..... III-3
III-2	MALDI-TOF of tryptic digest fragments of CS5-F and SC5-F ..... III-10
III-3	LCST analysis of CS5-F ..... III-11
III-4	Number of new N-termini created by HLE degradation of CS5 and CS5-F ..... III-13
III-5	Western-blot monitoring of disappearance of full-length CS5-F

upon treatment with HLE .....	III-14
III-6 HUVEC adhesion to engineered proteins .....	III-15
III-7 Kinetic analysis of HLE degradation .....	III-17
III-8 Control curves for Western blot densitometry.....	III-17
IV-1 Design and production of aECM-N <sub>3</sub> .....	IV-10
IV-2 Incorporation of <i>p</i> N <sub>3</sub> Phe as a function of concentration.....	IV-11
IV-3 FT-IR characterization of azide decomposition and mechanical properties of bulk films.....	IV-12
IV-4 Fluorescence microscopy and AFM characterization of films.....	IV-14
IV-5 Comparison of patterned proteins with mask .....	IV-15
IV-6 HUVEC attachment to photopatterned films.....	IV-17
V-1 Amino acid sequences of aECM proteins .....	V-2
V-2 Cloning strategy .....	V-4
V-3 DNA sequences of cell binding domains.....	V-5
V-4 DNA and protein sequence of ELF monomer .....	V-6
V-5. Agarose gel of multimerization reaction.....	V-9
V-6 Agarose gel of test digestions confirming ELF pentamer insert.....	V-10

## ABBREVIATIONS

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

<b>LCST</b>	lower critical solution temperature
<b>M9 or M9AA</b>	minimal medium
<b>MALDI-MS</b>	matrix-assisted laser desorption ionization-mass spectrometry
<b>mDHFR</b>	murine dihydrofolate reductase
<b>NEB</b>	New England Biolabs
<b>OD</b>	optical density
<b>PBS</b>	phosphate-buffered saline
<b>PEG</b>	poly(ethylene glycol)
<b>PheRS</b>	phenylalanyl-tRNA synthetase
<b>PMSF</b>	phenylmethylsulfonyl fluoride
<b>pN<sub>3</sub>Phe</b>	<i>para</i> -azidophenylalanine
<b>PP<sub>i</sub></b>	sodium pyrophosphate
<b>SDS-PAGE</b>	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
<b>Y-Ile</b>	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.*