PROTEIN ENGINEERING VIA SITE-SPECIFIC INCORPORATION OF NONNATURAL AMINO ACIDS

Thesis by

Inchan Kwon

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended January 19, 2007)

Acknowledgments

I would like to acknowledge my thesis advisor David Tirrell. I still remembered Dave's research presentation in my first year at Caltech. I was captivated by his clear presentation and fascinating research area, which influenced me to join his lab. During my days in his lab, I was deeply impressed by his endless curiosity in science and his systematic approach to solve problems with deep knowledge and strong enthusiasm. Without his timely advice and guidance throughout my PhD program, most of my projects would not have been successful. Dave was very supportive of my future career in all aspects.

I also would like to acknowledge other committee members, Prof. Frances Arnold, Prof. Dennis Dougherty, and Prof. Christina Smolke. They also served as my candidacy committee and provided me with many comments and much help since then. I appreciate their time and efforts to review my thesis as well as their great support for my career development.

I am grateful to all former and current Tirrell laboratory members. I was very lucky to work with such wonderful colleagues. In particular, Pin, Jamie, and Rebecca helped me a lot to settle down when I joined Tirrell laboratory. They knew everything I needed and were always willing to assist me. Three postdoctoral fellows, Kent, Jinsang, and Jin impressed me deeply with their commitment and enthusiasm to pursue science. They were also willing to share their experience and helped me to go through many non-scientific issues. They were invaluable role models as young scientists. I thank Caglar and Kimberly for their efforts and time reviewing my writing many times as well as for their helpful

discussions on diverse topics. I appreciate Julie and Jamie for sharing their experience in career development.

Finally, I wish to thank my wife Junghi and my son Hyungju for their continual support and encouragement throughout my PhD program. My daughter Suah, who was born recently, has added great pleasure to my life.

Abstract

Nonnatural amino acid incorporation has been one of the most important protein engineering techniques. Particularly site-specific incorporation of nonnatural amino acids would allow design of artificial proteins containing a nonnatural amino acid with minimal perturbation of their native properties. Site-specific incorporation of phenylalanine (Phe) analogs and tryptophan (Trp) analogs, such as *p*-bromophenylalanine (pBrF), *p*-iodophenylalanine, *p*-azidophenylalanine, 6-chlorotrytophan, 6-bromotryptophan, 5-bromotrytophan, and benzothienylalanine, into proteins in *Escherichia coli* has been realized by *E. coli* strains outfitted with a mutant yeast phenylalanyl-tRNA suppressor ($ytRNA^{Phe}_{CUA}$) and a mutant yeast phenylalanyl-tRNA synthetase (yPheRS (T415G)) with a point mutation in the active site of the enzyme. In order to reduce Trp and lysine (Lys) misincorporation at an amber codon, the $ytRNA^{Phe}_{CUA\_UG}$ containing the optimized sequence and the yPheRS (T415A) showing higher specificity toward pBrF were developed. Combining $ytRNA^{Phe}_{CUA\_UG}$ and yPheRS (T415A) allowed incorporation of pBrF into murine dihydrofolate reductase in response to an amber codon with at least 98% fidelity.

Re-assignment of degenerate sense codons offers the prospect of a substantially expanded genetic code and a correspondingly enriched set of building blocks for natural and artificial proteins. Here we describe the use of a mutant yeast phenylalanine transfer RNA ($ytRNA^{Phe}_{AAA}$) containing a modified anticodon to break the degeneracy of the genetic code in *E. coli*. By using an *E. coli* strain co-transformed with $ytRNA^{Phe}_{AAA}$ and a mutant yPheRS (T415G), we demonstrated efficient replacement of Phe by L-3-(2-naphthyl)alanine (2Nal) at UUU, but not at UUC codons.

However, this method had two limitations. First, the yPheRS (T415G) also activated Trp, which led to Trp misincorporation. Second, 2Nal was misincorporated at UUC Phe codons, due to the relaxed codon recognition of AAA anticodon in the $ytRNA^{Phe}_{AAA}$. High-throughput screening of a yPheRS library led to a more selective yPheRS variant (yPheRS_naph). The rationally designed $ytRNA^{Phe}_{CAA}$, which has the CAA anticodon recognizing only a UUG (Leu) codon, allowed incorporation of 2Nal only at UUG codon. Combined use of yPheRS_naph and $ytRNA^{Phe}_{CAA}$ achieved multiple-site-specific incorporation of 2Nal into proteins. These results illustrate a general method for increasing the number of distinct, genetically-encoded amino acids available for protein engineering.

Table of Contents

Acknowledgments iii

Abstract v

Table of Contents vii

List of Tables xii

List of Figures xiii

Chapter 1 Expansion of the Number of the Genetically Encoded Amino Acids in *E. coli* 1

Abstract 2
Introduction 2
Site-Specific Incorporation 3
Choice of a Nonnatural Amino Acid 4
Orthogonal Pairs 5
Modification of Endogenous Orthogonal Pairs 5
Orthogonal Pairs Derived from Yeast 6
Orthogonal Pairs Derived from Archaebacteria 7
Substrate Specificity Change 9
Rational Design of the Active Sites of Aminoacyl-tRNA Synthetases 9
Directed Evolution of Aminoacyl-tRNA Synthetase Substrate Specificity 12
New Codon-Anticodon Interactions 14
Nonsense Suppression 15
Breaking the Degeneracy of the Genetic Code 15
Frameshift Suppression 17
Conclusions 18
References 19

Chapter 2 Site-Specific Incorporation of Phenylalanine Analogs into Proteins in Vivo by an Engineered Yeast Phenylalanyl-tRNA Synthetase 36

Abstract 37
Introduction 38
Materials and Methods 40
Materials 40
Plasmid Construction for Synthetase Expression 40

Synthetase Expression and Purification 41
Amino Acid Activation Assay 42
Plasmid and Strain Construction for in Vivo Incorporation Assay 42
Analog Incorporation Assay in Vivo 43
Composition Analysis of Mutant mDHFR 44
Results and Discussion 45
Rationales for Engineering the New Synthetase Specificity 45
Synthetase Expression and Purification 46
Amino Acid Specificity of Mutant Yeast PheRS 47
Site-Specific Incorporation of Phe Analogs into mDHFR in Vivo Using a Phenylalanine Auxotrophic Strain 48
Site-specific Incorporation of Phe Analogs into mDHFR in Vivo Using a Phe/Trp Double Auxotrophic Strain 50
Site-specific Incorporation of Phe Analogs into mDHFR in Vivo Using a Phe/Trp/Lys Triple Auxotrophic Strain. 51
Conclusions 52
Acknowledgments 53
References 53

Chapter 3 Design of a Bacterial Host for Site-Specific Incorporation of *p*-Bromophenylalanine into Recombinant Proteins 68

Abstract 69
Introduction 69
Materials and Methods 71
Materials 71
Plasmid Construction for Synthetase Expression 72
Synthetase Expression and Purification 73
Amino Acid Activation Assay 73
Plasmid Construction for ytRNA^{Phe} Expression 74
In Vitro Transcription 75
Aminoacylation Assay 76
Strain and Plasmid Construction for in Vivo Incorporation of pBrF 77
In Vivo Incorporation of pBrF 77
Quantitative Analysis of pBrF Incorporation by Liquid Chromatography Mass Spectrometry (LC-MS) 78
Results and Discussion 79
Aminoacylation of ytRNA$^{Phe}_{CUA}$ and ytRNA$^{Phe}_{CUA_UG}$ with Lys 79
Occupancy of the Amber Site 80
Reduced Lys Misincorporation by ytRNA$^{Phe}_{CUA_UG}$ 81
Redesign of Phenylalanyl-tRNA Synthetase 81
ATP-PPi Exchange and Aminoacylation of ytRNA^{Phe} Catalyzed by yPheRS Variants 82
High Fidelity Site-Specific Incorporation of pBrF 83
Conclusions 83

Acknowledgments 84
References 84

Chapter 4 Site-specific Incorporation of Tryptophan Analogs into Recombinant Proteins in Vivo 96

Abstract 97
Introduction 98
Materials and Methods 102
Materials 102
Amino Acid Activation Assay 102
Strain and Plasmid Construction for in Vivo Incorporation Assays 103
In Vivo Analog Incorporation Assays 104
Quantitative Analysis of Analog Incorporation by Liquid Chromatography Mass Spectrometry (LC-MS) 105
Fluorescent Protein Isolation 105
Characterization of Spectral Properties 106
Results and Discussion 107
Rational Design of yPheRS and Choice of Trp Analogs 107
Activation of Trp Analogs by yPheRS (T415G) 108
Site-Specific Incorporation of Trp Analogs into mDHFR in Vivo 109
Spectral Analysis of Fluorescent Proteins 111
Conclusions 113
Acknowledgments 113
References 114

Chapter 5 Breaking the Degeneracy of the Genetic Code 127

Abstract 128
Introduction 128
Materials and Methods 129
Materials 129
Strains and Plasmids 129
Construction of Plasmid Carrying Mutant Yeast tRNA 129
Construction of Plasmid Carrying mDHFR and Mutant Yeast Phenylalanyl-tRNA Synthetase 130
Protein Expression and Purification 130
Amino Acid Analysis 131
MALDI-MS Analysis 131
Results and Discussion 132
Conclusions 137
Acknowledgments 137
References 137

Chapter 6 Breaking the Degeneracy of the Leucine Codons in *Escherichia coli* 146

Abstract 147
Introduction 147
Materials and Methods 151
Materials 151
Preparation of *E. coli* hosts 151
Plasmid Construction for Reporter Gene Expression 151
Construction of yPheRS Library 153
Construction of Expression Library 154
Screening of yPheRS Library 154
Characterization of the Isolated Clones 155
Amino Acid Activation Assay 156
Construction of Plasmids and Expression Hosts for Incorporation of 2Nal at Phe Codons 156
Residue- and Site-Specific Incorporation of 2Nal into Recombinant Proteins 158
Construction of Plasmids and Expression Hosts for 2Nal Incorporation at Leu Codons 159
Expression of mDHFR Variants and GFP Variants in Vivo 160
Quantitative Analysis of Codon Occupancy 161
Results and Discussion 162
Misincorporation of Trp into Recombinant Proteins in Response to UUU Codons 162
Construction of yPheRS Library 163
High-throughput Screening of yPheRS library 163
Characterization of the Selective yPheRS Variant 164
Amino Acid Activation Analysis by yPheRS_naph 165
Elimination of Trp Misincorporation at UUU Codons by yPheRS_naph in Vivo 166
Residue- and Single-Site-Specific Incorporation of 2Nal in Vivo 167
Single-Site-Specific Incorporation of 2Nal into GFP in Vivo 167
Misincorporation of 2Nal at Unwanted Sites (UUC Codons) in Vivo 168
Breaking the Degeneracy of the Leucine Codons 169
UUG Codon-Specific Incorporation of 2Nal 170
Minimal Perturbation of Native Properties of GFP upon Incorporation of 2Nal 171
Conclusions 171
Acknowledgments 172
References 172

Appendix A Plasmid Information 189

Appendix B Marker Protein Information 209

Appendix C Names and Sequences of PCR Primers 215

List of Tables **Page**

Table 2.1	Kinetic parameters for ATP-PPi exchange of amino acids (**1**-**7**) by the mutant yeast PheRS	61
Table 3.1	Occupancy of amber sites and expression yields	93
Table 3.2	Kinetic parameters for ATP-PPi exchange of amino acids by the wild-type yPheRS, yPheRS (T415G), and yPheRS (T415A) variant	94
Table 4.1	Kinetic parameters for ATP-PPi exchange by the yPheRS (T415G)	124
Table 4.2	Occupancy of amber sites and mDHFR expression yields	125
Table 4.3	Spectral properties of CFP6 variants	126
Table 5.1	Molar masses and numbers of 2Nal residues observed for mDHFR samples prepared under various conditions	140
Table 6.1	Kinetic parameters for ATP-PPi exchange of amino acids by the yPheRS (T415G) and yPheRS_naph variants	187
Table 6.2	Occupancy of UUU and UAG codons by various amino acids	188

List of Figures **Page**

Figure 1.1 Nonnatural amino acids that have been genetically incorporated into proteins in bacteria, yeast or mammalian cells. 28

Figure 1.2 Structures of yeast $tRNA^{Phe}$ and *E. coli* $tRNA^{Phe}$. 29

Figure 1.3 Anticodon-loop tRNA library, all-loop tRNA library derived from *M. jannaschii* $tRNA^{Tyr}_{CUA}$ and a general selection scheme for suppressor tRNAs that are orthogonal to *E. coli* endogenous systems and charged efficiently by a cognate synthetase. 30

Figure 1.4 Crystal structure of the *T. thermophilus* PheRS (*t*PheRS, pdb 1B70) in ribbon model and sequence alignment of PheRS from 21 different organisms. 31

Figure 1.5 A Phe substrate and four residues within 7 Å of the *para*-position of phenyl ring of a Phe substrate inside the binding pocket of a homology model of yPheRS structure, and a screening scheme of yPheRS library. 33

Figure 1.6 A schematic diagram describing the concept of breaking the degeneracy of the genetic code. 35

Figure 2.1 Sequence alignment of PheRS variants from *Thermus thermophilus, Escherichia coli, Saccharomyces cerevisiae* and stereoview of active site of PheRS from *T. thermophilus*. 59

Figure 2.2 Amino acids involved in this study. (**1**) phenylalanine; (2) *p*-bromophenylalanine; (**3**) *p*-iodophenylalanine; (**4**) *p*-azidophenylalanine; (**5**) tyrosine; (**6**) trytophan; (**7**) 3-(2-naphthyl)alanine; (**8**) L-lysine. 60

Figure 2.3 Amino acid sequence of marker protein mDHFR (38Am). 62

Figure 2.4 MALDI-MS of proteolytic peptide fragments derived from mDHFR (38Am) expressed in media supplemented with **7** or **2.** 63

Figure 2.5 The tandem mass spectrum of Peptide A (NGDLPWPPLRNE***Z***K) derived from mDHFR (38Am) expressed in AF strain in media supplemented with **7.** 65

Figure 2.6 MALDI-MS of proteolytic peptide fragments of mDHFR (38Am) expressed in AFWK strain in media supplemented with **2** or **3.** 66

Figure 3.1	Aminoacylation of ytRNA$^{Phe}_{CUA}$ and ytRNA$^{Phe}_{CUA_UG}$ with Lys by eLysRS.	88
Figure 3.2	LC-MS chromatograms of tryptic digests of mDHFR.	89
Figure 3.3	Charging of Phe and Trp by wild-type yPheRS, yPheRS (T415G), and yPheRS (T415A).	91
Figure 4.1	Amino acids involved in this study: (**1**) tryptophan, (**2**) phenylalanine, (**3**) 6-chlorotryptophan, (**4**) 6-bromotryptophan, (**5**) 5-bromotryptophan, (**6**) benzothienylalanine.	118
Figure 4.2	A hypothetical model for the adaptability of Trp analogs to the binding pocket of eTrpRS.	119
Figure 4.3	SDS-PAGE analysis of mDHFR_38Am expression.	120
Figure 4.4	Tandem mass spectrum of the peptide (NGDLPWPPLRNEZK).	121
Figure 4.5	Crystal structure of enhanced cyan fluorescent protein and chromophores of CFP6 variants containing **1**, **3**, **4**, and **6**.	122
Figure 4.6	Absorption and fluorescence emission spectra for CFP6 variants containing different Trp analogs at the Trp66 position.	123
Figure 5.1	A strategy for multi-site incorporation of 2Nal into recombinant proteins by breaking the degeneracy of the phenylalanine codons.	141
Figure 5.2	SDS-PAGE analysis of mDHFR prepared in minimal media supplemented with 3 mM 2Nal and free of exogenous Phe.	142
Figure 5.3	Replacement of Phe by 2Nal can be detected in MALDI-MS spectra of tryptic fragments of mDHFR.	143
Figure 5.4	Tandem mass spectrum of Peptide 1_{UUU} (Nal) YKF*EVYEK.	145
Figure 6.1	LC-MS chromatogram of tryptic digests of mDHFR.	176
Figure 6.2	Phe substrate and four residues within 7 Å of *para*-position of the phenyl ring of the substrate inside the binding pocket of a homology model of yPheRS structure.	178
Figure 6.3	A screening scheme for yPheRS library.	179

Figure 6.4	Fluorescence intensities of cells containing GFP6 expressed in minimal medium without or with 2Nal.	180
Figure 6.5	Fluorescence intensities of cells under uninduced conditions.	181
Figure 6.6	UUC and UUU codon occupancy by Phe and 2Nal.	182
Figure 6.7	Fluorescence intensities of cells expressing GFP3_WC and GFP3 expressed in minimal medium without or with 2Nal.	183
Figure 6.8	LC-MS chromatogram of tryptic digests of mDHFR.	184
Figure 6.9	Fluorescence intensities of cells containing GFP3 (158UUG) expressed in minimal medium without or with 2Nal.	186