

**Expanding the Biosynthetic Capacity of the Aminoacyl-tRNA  
Synthetases**

Thesis by

Pin Wang

In Partial Fulfillment of the Requirements for Degree of Doctor of  
Philosophy

California Institute of Technology

Pasadena, California, USA

2003

(Defended on June 3<sup>rd</sup>, 2003)

© 2003

Pin Wang

All Rights Reserved

To Lili

To My Parents and My Sister

## ACKNOWLEDGEMENTS

Before I started to work on my thesis, I thought acknowledgements section would be the easiest one. In reality I found that it was far more difficult than I expected to put together this part of thesis, primarily due to the fact that the bulk of this work would not be completed without the influence of many people in my life. I thank everyone who encouraged me and supported me in many aspects of my life. It is almost impossible to list all names herein and I am sure I have missed some names, for which I apologize.

Foremost, I would like to acknowledge my thesis advisor, David Tirrell, for having guided me into this fascinating area of research, for his constant support and encouragement, for sharing with me his ethological perspectives on science, and for teaching me to follow people's opinion as long as "the opinion does not kill you." I admire his energy, his enthusiasm for science and his good memory. Dave allowed me to work on several projects and to pursue collaborations with different groups. The opportunities I have been afforded in his group were enormous. Dave is extremely supportive of my future career development, has given me confidence to look forward, and helps me in every possible aspect. I am looking forward to working with him again in the near future.

I also would like to thank other members of my committee, Prof. Mark Davis, Prof. Bill Goddard, Prof. Steve Mayo, for their contributions to this work. Their patience and insights have helped move forward much of this thesis work. I appreciate their time dedicated to reviewing my thesis results. In particular, the collaborative efforts on computational design of the aminoacyl-tRNA synthetases with Bill and Steve constitute the critical components of this thesis. Their invaluable suggestions came at a different

angle, which helped me a lot on design of experiments and taught me the computational knowledge. The interactions with Vaidehi Nagarajan from Bill's group and Deepshikha Datta from Steve's group are very enjoyable and valuable. Vaidehi mentored me on computational methods and was patient to answer every of my naïve questions.

I would like to take this opportunity to thank Prof. Zhen-Gang Wang for recruiting me to Caltech. My life at Caltech began with his firm trust and encouragements. Prof. Julie Kornfield was kind enough to allow me to work in her lab for the first summer while Dave was still at UMass. I am grateful for the support from Bill, Steve, Prof. Francis Arnold and Prof. Julie Kornfield during my academic job hunting.

The experience with the entire Tirrell group was wonderful and unforgettable. My progress and accomplishments are always associated with their assistance and ideas. When I first joined the group, Jeff, Kristi, Wendy, Karen, Dave, Jill, Kathy and Nandita from UMass had me settle down in the laboratory. I am very fortunate to have worked with other former members that passed through Caltech during the last five years, particularly Yi, Kent, Ralf, Kevin, Jens, Thierry, Hanna, Youngback, Toshi, Angelika. I especially appreciate the time I spent working with my good colleague and friend Yi Tang. He taught me the essential skills to carry out most of the thesis work. The experience with him was stimulating and inspiring. His never satisfactions of the accomplished work really influenced me and motivated me. His advice for my academic job search was extremely helpful. I am looking for the opportunity to work with him in the near future. I thank Kent for the rigorous and deep discussion on science and encouragements regarding job search. I also would like to thank the current members of

the group, including Isaac, Sarah, Marrison, Paul, Jinsang, Wei, Jamie, Rebecca, Charles, Julie, Mike, Soojin, Chris, Jin, Kimberley and many others for their friendliness and help. Isaac provided excellent advice on my organic chemistry. Sarah and Jamie helped me with writing English. Jinsang shared with me a lot of his experience in science and pushed me with my job application. Paul made our office more enjoyable. Qualifying exam with Wei was unforgettable. Working with Soojin was also wonderful. I am very fortunate to have this group of colleagues and am looking forward to working with them again in the future.

Life outside the laboratory at Caltech is truly wonderful. The experience with Yi and Dave when we won the 10K business competition is one of the most exciting moments at Caltech. I am really lucky to have interacted with Prof. John Baldeschwieler and taken his E102 class. His enthusiasm about entrepreneurship made our hypothetical company, Duragene, almost real. Together with Chunhui Mo, Joe Andrieu, Girish N. Aakalu, Lili Yang, Ying Go, we founded Caltech Consulting Club. Knowledge gleaned from this club and experiences obtained from two summer projects with two different companies are precious.

I have been very fortunate to gain many lasting friendships over the last five years at Caltech. These friends have been an important part of my graduate school experience not only on a scientific basis but on a personal level as well. I would like to thank the gang like Chunhui, Hongyu, Qiang, Ying, Lan, Huazhang, Xianglei, Wei, Weijun, Lin, Junhua, and Jay for all the good times. Thanks also go to my friends in Boston like Ying, Minghao, Jian and old friends like Xianzhi, Lixin. It is Minghao who encouraged me to transfer to Caltech.

I would like to thank my parents for their love and care. Also I would like to thank my uncle Y. C. Yang and his family for helping me study and live in the US.

Last but definitely not least, I thank my wonderful girlfriend Lili Yang. My biggest achievement at Caltech is having her in my life. Her unconditional support, kindness, and love completely changed my life in the last three years. She made my life colorful and fruitful. I am truly thanking her for everything she has done for me.

The research in this thesis was funded by NIH Grant R01-GM62523 and by the NSF Center for the Science and Engineering of Materials at Caltech.

**ABSTRACT**

Incorporation of non-natural amino acids into proteins *in vivo* can provide biological materials with new chemical functions and improved physical properties. Examples include new posttranslational modification chemistry by introducing azide and ketone moieties into recombinant proteins, and novel strategies for engineering hyperstable proteins by incorporating fluorinated side chains. Implementing such methods requires manipulation of protein biosynthesis to specifically alter the genetic code. The rules of the genetic code are established by the aminoacylation reaction, where the aminoacyl-tRNA synthetases (aaRS) catalyze the attachment of the amino acids to their cognate tRNAs. Thus manipulation of cellular aminoacylation reactions could potentially expand the available set of amino acid building blocks for protein engineering and biomaterials engineering.

By simple depletion of the cellular pool of isoleucine and utilization of isoleucine auxotrophic hosts, we were able to force the endogenous isoleucyl-tRNA synthetase to join 5,5,5-trifluoroisoleucine (5TFI) to tRNA<sup>Ile</sup> and assign it to isoleucine codons *in vivo*. Murine interleukin 2 containing 5TFI retains its biological activity. We showed that engineering of bacterial expression hosts can allow a single RNA message to be read in different ways, depending on the relative rates of competing aminoacylation reactions. Specifically, we showed that the 2*S*,3*R*-form of 4,4,4-trifluorovaline can be assigned either to isoleucine or to valine codons, depending on whether the bacterial host overexpresses the isoleucyl- or the valyl-tRNA synthetase. When an amino acid analog of interest is not recognized by the corresponding wild-type aaRS, we can either identify the appropriate modification of the amino acid as a promising ligand or design new



synthetase activity. We describe an attempt to develop a virtual ligand screening method to find non-natural amino acids that can serve as ligands for the phenylalanyl-tRNA synthetase and our computational results correlate well with experimental results *in vitro* and *in vivo*. We also present a computational method for identifying the sites of mutations to relax the substrate specificity of the *E. coli* phenylalanyl-tRNA synthetase (*ePheRS*). One designed variant of *ePheRS* allows the efficient *in vivo* incorporation of aryl ketone functionality into proteins *in vivo*. Proteins outfitted with ketone functionality can be chemoselectively ligated with hydrazide reagents under mild conditions. Three designed mutants of *ePheRS* were subjected to extensive examination, and a broad activation profile toward many non-natural aromatic amino acids was observed. *E. coli* host strains were established to over-express these mutant *ePheRS*s, enabling the re-assignment of the Phe codons to many non-natural amino acids. By rational attenuation of the editing function of a leucyl-tRNA synthetase, oxonorvaline was incorporated into a recombinant protein in *Escherichia coli*.

The work described above addresses the multi-site incorporation of new amino acids into proteins *in vivo*, which can be utilized to engineer the overall properties of biomacromolecules such as protein stability. The second component of this thesis focused on the site-specific incorporation of novel amino acids into proteins *in vivo*, which can be applied to problems that require local change of protein behavior. We have refined a previously described system, where we introduce a mutant form of yeast PheRS co-transformed with a cognate suppressor tRNA, allowing incorporation of several aromatic amino acids into proteins in response to an amber codon. The results firmly

demonstrate the general strategy of importing an exogenous synthetase/tRNA pair to achieve site-specific incorporation of non-natural amino acids into proteins *in vivo*.

## TABLE OF CONTENTS

	Page
<b>ACKNOWLEDGEMENTS</b>	iv
<b>ABSTRACT</b>	viii
<b>LIST OF TABLES</b>	xv
<b>LIST OF FIGURES</b>	xvi
<b>CHAPTER</b>	
<b>Chapter 1 Introduction to Aminoacyl-tRNA Synthetases and Incorporation of Non-natural Amino Acids</b>	1-1
1. Aminoacyl-tRNA Synthetases (aaRSs)	1-1
2. Isoleucyl-tRNA Synthetase (IleRS)	1-5
3. Valyl-tRNA Synthetase (ValRS)	1-8
4. Leucyl-tRNA Synthetase (LeuRS)	1-10
5. Phenylalanyl-tRNA Synthetase (PheRS)	1-14
6. Site-specific Incorporation of Non-natural Amino Acids into Proteins by Chemically Misacylated tRNA	1-16
7. Site-specific Incorporation of Non-natural Amino Acids into Proteins <i>In Vivo</i>	1-20
8. Residue-specific Incorporation of Non-natural Amino Acids into Proteins <i>In Vivo</i>	1-22
9. References	1-55
<b>Chapter 2 Virtual Screening for Binding of Phenylalanine Analogs to Phenylalanyl-tRNA Synthetase</b>	2-1
1. Introduction	2-2
2. Methods	2-4
2.1. Computational Methodology	2-4
2.2. Procedures for Screening Analogs and Application to PheRS	2-7
2.3. Measurement of Relative Free Energies	2-11
3. Results and Discussion	2-12
3.1. Prediction of the Binding Site of Phe in PheRS	2-12
3.2. Calculation of Binding Energies for Phe Analogs	2-13
3.3. Predicted Binding Sites for p-Fluorophenylalanine and 3-Thienylalanine	2-15
4. Conclusion	2-16
5. References and Notes	2-38
<b>Chapter 3 Incorporation of Trifluoroisoleucine into Proteins <i>In Vivo</i></b>	3-1

1.	Introduction	3-2
2.	Materials and Methods	3-3
2.1.	Synthesis of Amino Acid Analogs	3-4
2.2.	Determination of Translational Activity	3-8
2.3.	Protein Expression and Incorporation of Trifluoro- isoleucine into mDHFR	3-9
2.4.	Protein Purification	3-10
2.5.	Tryptic Peptide Analysis	3-10
2.6.	IleRS Cloning, Expression and Purification	3-10
2.7.	ATP-PP <sub>i</sub> Exchange Assay	3-11
2.8.	Incorporation of 5TFI into Murine Interleukin-2	3-12
3.	Results	3-13
3.1.	Synthesis of Isoleucine Analogs	3-13
3.2.	Incorporation of 5TFI into Protein <i>In Vivo</i>	3-14
3.3.	Effect of 5TFI on Cell Growth	3-15
3.4.	Amino Acid Activation <i>In Vitro</i>	3-16
3.5.	5TFI in Murine Interleukin-2	3-16
4.	Discussion	3-17
5.	References	3-38
<b>Chapter 4</b>	<b>It Depends on How You Read It: Alternative Translations of a Single RNA Message</b>	4-1
1.	Introduction	4-2
2.	Materials and Methods	4-2
2.1.	Materials	4-2
2.2.	Synthetase Cloning, Expression and Purification	4-3
2.3.	Amino Acid Activation Assays	4-4
2.4.	Plasmid Construction for <i>In Vivo</i> Assays of Codon Assignment	4-4
2.5.	<i>In Vivo</i> Translational Assays of Codon Assignment	4-5
2.6.	Amino Acid Identification at Valine and Isoleucine Codon Sites	4-6
3.	Results and Discussion	4-7
4.	Conclusion	4-11
5.	References and Notes	4-23
<b>Chapter 5</b>	<b>A Designed Phenylalanyl-tRNA Synthetase Variant Allows Efficient <i>In Vivo</i> Incorporation of Aryl Ketone Functionality into Proteins</b>	5-1
1.	Introduction	5-2
2.	Materials and Methods	5-3
2.1.	Plasmid Construction	5-3
2.2.	Determination of Translational Activity	5-4
2.3.	Protein Purification	5-5

	2.4.	Tryptic Peptide Analysis	5-5
	2.5.	Chemical Modification of Protein	5-5
3.		Results and Discussion	5-6
4.		Conclusion	5-9
5.		References and Notes	5-22
<b>Chapter 6</b>		<b>Engineering a Relaxed Substrate Specificity of <i>E. coli</i> Phenylalanyl-tRNA Synthetase to Incorporate a Diverse Set of Amino Acid Analogs into Proteins <i>In Vivo</i></b>	6-1
	1.	Introduction	6-2
	2.	Materials and Methods	6-5
	2.1.	Materials	6-5
	2.2.	Computational Methodology and Its Design of Mutant Forms of PheRS	6-5
	2.3.	Plasmid Construction for Synthetase Expression	6-7
	2.4.	Synthetase Expression and Purification	6-8
	2.5.	Amino Acid Activation Assays	6-9
	2.6.	Aminoacylation Assays	6-9
	2.7.	Plasmid Construction for <i>In Vivo</i> Incorporation Assays	6-10
	2.8.	Analog Incorporation Assays <i>In Vivo</i>	6-11
	2.9.	Target Protein Composition Analysis	6-12
	3.	Results and Discussion	6-12
	3.1.	Computational design	6-12
	3.2.	Expression and Purification of ePheRS Variants	6-14
	3.3.	Activation of Analogs by Variant Enzymes <i>In Vitro</i>	6-15
	3.4.	Aminoacylation by Variant Enzymes <i>In Vitro</i>	6-17
	3.5.	<i>In Vivo</i> Effects of Mutant Synthetases	6-18
	3.6.	Protein Tryptic Peptide Analysis	6-20
	4.	Conclusion	6-23
	5.	References	6-42
<b>Chapter 7</b>		<b>Introduction of an Aliphatic Ketone into Recombinant Proteins in a Bacterial Strain that Overexpresses an Editing-impaired Leucyl-tRNA Synthetase</b>	7-1
	1.	Introduction	7-2
	2.	Materials and Methods	7-2
	2.1.	Materials	7-2
	2.2.	Synthetase Expression and Purification	7-3
	2.3.	ATP-PP <sub>i</sub> Exchange Assay	7-3
	2.4.	Expression Plasmids	7-4
	2.5.	Analog Incorporation Assay	7-4
	2.6.	Protein Composition Analysis	7-5
	2.7.	Chemical Modification of Ketone in Protein	7-6

3.	Results and Discussion	7-6
4.	Conclusion	7-9
5.	References	7-18
<b>Chapter 8</b>	<b>Site-specific Incorporation of Amino Acid Analogs into Proteins <i>In Vivo</i> by an Engineered Yeast Phenylalanyl-tRNA</b>	<b>8-1</b>
1.	Introduction	8-2
2.	Materials and Methods	8-4
2.1.	Materials	8-4
2.2.	Plasmid Construction for Synthetase Expression	8-4
2.3.	Synthetase Expression and Purification	8-5
2.4.	Amino Acid Activation Assays	8-6
2.5.	Plasmid Construction for <i>In Vivo</i> Incorporation Assays	8-6
2.6.	Analog Incorporation Assays <i>In Vivo</i>	8-7
2.7.	Composition Analysis of the Mutant mDHFR	8-8
3.	Results and Discussion	8-9
3.1.	Rationales for Engineering the New Synthetase Specificity	8-9
3.2.	Synthetase Expression and Purification	8-10
3.3.	Amino Acid Specificity of the Mutant Yeast PheRS	8-11
3.4.	Site-specific Incorporation of Analogs into mDHFR <i>In Vivo</i> Using Phenylalanine Auxotrophic Strain	8-12
4.	Conclusion	8-16
5.	References	8-32

## LIST OF TABLES

		Page
Table 1-1	Classes of aminoacyl-tRNA synthetases	1-27
Table 2-1A	Comparison of various predicted and experimental structures for PheRS	2-20
Table 2-1B	Notation used to denote the different ligand protein complex	2-20
Table 2-2A	Hydrogen bond distances and vdW interactions in the Phe binding sites of PheRS/Phe complexes	2-21
Table 2-2B	All residues in van der Waals contact of Phe in the Phe/PheRS structure	2-21
Table 2-3	Binding energy and its components for Phe analogs calculated from HierDock	2-22
Table 2-4A	Hydrogen bond distances of Phe, <i>p</i> -fluoro-phenylalanine and 3-thienylalanine analogs in their respective binding sites	2-23
Table 2-4B	List of van der Waals contacts of the side chains of <i>p</i> -fluoro-phenylalanine and 3-thienylalanine	2-23
Table 3-1	Protein yield and extent of isoleucine replacement in mDHFR	3-20
Table 3-2	Kinetic parameters for activation by <i>E. coli</i> IleRS	3-21
Table 3-3	Proliferative response of H2-T cells	3-22
Table 4-1	Kinetic parameters for activation of amino acids by <i>E. coli</i> IleRS and ValRS	4-13
Table 6-1	ORBIT calculation for <i>p</i> -acetyl-phenylalanine binding into <i>t</i> PheRS	6-27
Table 6-2	ATP-PPi exchange kinetics of wild-type and mutant forms of <i>e</i> PheRS toward canonical and non-canonical amino acids	6-28
Table 6-3	Mass data for peptide fragment 1 derived from mDHFR	6-29
Table 7-1	Amino acid activation by wild-type LeuRS	7-11
Table 8-1	The kinetic parameters for the ATP-PPi exchange of amino acids by the yeast mutant PheRS	8-19

## LIST OF FIGURES

	Page	
Figure 1-1	The role of aminoacyl-tRNA synthetase in the DNA-templated protein biosynthesis	1-28
Figure 1-2	Structure of tRNA and its interaction with aaRS	1-30
Figure 1-3	Crystal structure of IleRS from <i>T. thermophilus</i>	1-32
Figure 1-4	Cognate isoleucine and non-cognate valine engulfed in the synthetic site of IleRS from <i>T. thermophilus</i>	1-34
Figure 1-5	Stereoview of Ile-AMS bound to the aminoacylation site of IleRS from <i>T. thermophilus</i>	1-36
Figure 1-6	tRNA bound to IleRS from <i>S. aureus</i>	1-38
Figure 1-7	Crystal structure of ValRS complexed with tRNA <sup>Val</sup> from <i>T. thermophilus</i>	1-40
Figure 1-8	Aminoacyl adenylate analogue Val-AMS bound to the active site of ValRS	1-42
Figure 1-9	Crystal structure of LeuRS from <i>T. thermophilus</i>	1-44
Figure 1-10	Active site of LeuRS from <i>T. thermophilus</i> complexed with Leu-AMS	1-46
Figure 1-11	Overall crystal structure of PheRS and $\alpha$ -subunit bound with substrate from <i>T. thermophilus</i>	1-48
Figure 1-12	Active site of PheRS from <i>T. thermophilus</i> complexed with PheOH-AMP	1-50
Figure 1-13	Site-specific incorporation of non-natural amino acids into proteins through chemically misacylated tRNA and nonsense suppression	1-52
Figure 2-1	Comparison of crystal structure without ligand in red with the forcefield minimized structure in yellow.	2-24
Figure 2-2	$\alpha$ -subunit of <i>T. thermophilus</i> PheRS from the crystal structure	2-26
Figure 2-3	Comparison of binding pocket for PheRS	2-28
Figure 2-4	Predicted binding site of Phe in PheRS	2-30
Figure 2-5	Calculated binding energies for analogs of Phe in PheRS	2-32
Figure 2-6	The van der Waals surface of <i>p</i> -bromo-phenylalanine clashing with the vdW surface of the side chain of Ala 314	2-34
Figure 2-7	Comparison of binding pocket for phenylalanine to that of <i>p</i> -fluorophenylalanine and 3-thienylalanine	2-36
Figure 3-1	SDS-PAGE analysis of mDHFR synthesized by AF-IQ[pQE15] or AF-IQ[pQE-IleRS]	3-26
Figure 3-2	MALDI analysis of mDHFR following trypsin digestion	3-28
Figure 3-3	Growth rate of <i>E. coli</i> strain AF-IQ[pQE15] as a function of the concentration of 5TFI in the culture medium	3-30
Figure 3-4	MALDI analysis of mIL-2 following trypsin digestion	3-32
Figure 3-5	Proliferative response of IL-2 dependent H2-T cells to recombinant mIL-2 proteins	3-34
Figure 3-6	Crystal structure of hIL-2	3-36



Figure 4-1	SDS-PAGE analysis for <i>in vivo</i> incorporation of non-canonical amino acids into recombinant protein mDHFR	4-14
Figure 4-2	New assignment of isoleucine and valine codons can be detected by MALDI-MS analysis of tryptic fragment of mDHFR	4-16
Figure 4-3	The tandem mass spectrum of the peptide	4-18
Figure 4-4	The tandem mass spectrum of the peptide	4-20
Figure 5-1	Active sites in <i>t</i> PheRS/Phe	5-12
Figure 5-2	SDS-PAGE of cell lysates of 4 hr post-induction with 1 mM IPTG	5-14
Figure 5-3	Matrix-assisted laser desorption ionization mass spectrometry of purified proteins	5-16
Figure 5-4	MALDI TOF mass spectra of tryptic peptides digested from mDHFR	5-18
Figure 5-5	Western blot showing chemoselective modification of ketone functionality in mDHFR	5-20
Figure 6-1	Overall crystal structure of the <i>T. thermophilus</i> PheRS and sequence alignment of PheRS from different organisms	6-30
Figure 6-2	Chemical structure of amino acids involved in this study	6-32
Figure 6-3	Ribbon representation of the portion of catalytic $\alpha$ -subunit of PheRS from <i>T. thermophilus</i>	6-34
Figure 6-4	Aminoacylation of tRNA <sup>Phe</sup> by wild-type and mutant <i>e</i> PheRS	6-36
Figure 6-5	Amino acid sequence of target protein mDHFR	6-38
Figure 6-6	MALDI-MS of tryptic peptide fragment 2 derived from mDHFR	6-40
Figure 7-1	Incorporation of 2 and 3 into target protein A1 and MALDI-MS of protein shown in lane 8	7-12
Figure 7-2	Chemoselective modification of the ketone moiety in target protein A1	7-14
Figure 7-3	Incorporation of unnatural amino acids by different LeuRS editing mutants	7-16
Figure 8-1	Sequence alignment of PheRS variants and stereoview of active site of PheRS from <i>T. thermophilus</i> .	8-20
Figure 8-2	Amino acids involved in this study	8-22
Figure 8-3	Activation of amino acids (1-12) by yeast mutant PheRS	8-24
Figure 8-4	Amino acid sequence of marker protein mDHFR	8-26
Figure 8-5	MALDI-MS of proteolytic peptide fragments derived from mDHFR	8-28
Figure 8-6	The tandem mass spectrum of Peptide A	8-30

