Characterization of the \textit{in vitro} and \textit{in vivo} specificity of \textit{trans}-editing proteins and interacting aminoacyl-tRNA synthetases

\textbf{DISSERTATION}

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Ziwei Liu, B.S.

Graduate Program in Biochemistry

The Ohio State University

2014

Dissertation Committee:

Karin Musier-Forsyth, Advisor

Juan Alfonzo

Charles Bell

Kotaro Nakanishi
Abstract

Aminoacyl-tRNA synthetases (ARSs) are responsible for decoding genetic information, whereby each amino acid (aa) is attached to a cognate tRNA. The accuracy of ARSs is critical for viable cellular functions. Errors arise from the misactivation of noncognate amino acids in the synthetic sites of ARSs and the misacylation of these substrates onto specific tRNAs. Editing of the mischarged tRNAs may be provided by the ARSs themselves in cis and by freestanding proteins that are homologs to the editing domain of ARSs in trans. A family of editing factors that are homologs of the bacterial prolyl-tRNA synthetase (ProRS) editing domain, collectively known as the “INS superfamily”, is encoded in organisms from bacteria to humans. This family includes the cis-editing domain (INS) of ProRS, which clears Ala-tRNA^Pro, the trans-editing factor YbaK, which clears Cys-tRNA^Pro, and 4 additional single-domain homologs (ProXps). In this work, we investigate YbaK, as well as ProXp-y and ProXp-z, INS homologs of unknown function.

An in vivo screen of 230 conditions wherein an Escherichia coli ProXp-y deletion strain was grown in the presence of elevated levels of amino acids and specific ARSs, together with the results of in vitro deacylation assays, reveal Ser/Thr-tRNA deacylase function for this homolog. These results, along with in vitro studies of Bordetella parapertussis suggest that these free-standing editing domains have the capability to
prevent mistranslation errors caused by a number of ARSs, including LysRS, ThrRS and AlaRS. The expression of these domains is likely to provide a selective growth advantage to organisms subjected to environmental stresses and other conditions that alter the amino acid pool. The in vivo role of YbaK was tested under oxidative stress conditions. An increase in the lag phase of the \( ybaK \) null strain under elevated \( \text{H}_2\text{O}_2 \) concentration relative to the wild type strain supports the hypothesis that YbaK helps bacteria defend against reactive oxygen species.

Recent studies have shown that in addition to the proteinogenic amino acids, there are many noncanonical amino acids or metabolite intermediates that can be mischarged by ARSs and misincorporated into nascent peptides. We explored several analogues that are similar in structure or shape to Ser and Thr, and tested the possibility that these analogues can be mischarged by various ARSs and deacylated by ProXp-y. Several new analogues were found to be substrate for ARSs and ProXp-y \textit{in vitro}. In addition, studies were initiated to discover interacting partners of ProXp-y \textit{in vivo}. Metabolomic and crystallization studies were also initiated to understand the amino acid pool under different conditions and the interactions of ProXp-y with its partners. These studies could provide more information on the substrate specificity of ProXp-y and expand our knowledge of the potential role of freestanding proteins in living cells.
Dedication

This document is dedicated to my family.
Acknowledgements

First, I am deeply grateful to my PhD advisor, Dr. Karin Musier-Forsyth, for her everlasting support and encouragement over the years. Without her insightful advice and unusual patience, I could not be able to overcome all the difficulties in the research and understand the meaning of science, and could never grow as a determined scientist. I have learned not only the way to do experiment, but also the importance of working and interacting with others in the scientific community. I will always remember the time she inspired me when I met failures and the moments when we shared good progresses. I sincerely appreciate the opportunities she gave me to learn and enjoy the science from different collaborations and meetings.

Secondly, I want to thank my committee members and other collaborators during my PhD research. I appreciate Professor Michael Ibba for giving me numerous valuable suggestions on my projects. Professor Kurt Fredrick, Professor Irina Artsimovitch and Professor Kotaro Nakanishi gave insightful comments in the journal club. Professor Charles Bell helped me a lot on X-ray crystallography study and Professor Juan Alfonzo gave serious advices in my committee meeting.

I would also like to thank all my friends and colleagues in the group, who help me in the work and support me in the life. I appreciate the advices they gave me during the hard time, and I always enjoy the busy but fun days working together.
Finally, I want to thank my family for their understanding and love. Though separated due to my study, they always encourage me to move on and be brave.
Vita

2008.................................................................B.S. Biology, Tsinghua University, China

2008 to present .................................................Graduate Research Assistant, Department of Chemistry and Biochemistry, The Ohio State University

Fields of Study

Major Field: Biochemistry
Table of Contents

Abstract .............................................................................................................................. ii

Dedication ........................................................................................................................ iv

Acknowledgement ........................................................................................................... v

Vita ..................................................................................................................................... vii

Fields of Study .................................................................................................................. vii

Table of Contents ............................................................................................................ viii

List of Tables .................................................................................................................... xii

List of Figures ................................................................................................................... xiii

Chapter 1: Introduction ..................................................................................................... 1

1.1 Fidelity of translation ................................................................................................. 1

1.2 Aminoacyl-tRNA synthetases .................................................................................. 1

1.2.1 Aminoacylation reaction and ARSs class system .................................................. 1

1.2.2 Specificity of ARSs .............................................................................................. 5

1.3 Editing by aminoacyl-tRNA synthetases .................................................................. 8

1.3.1 Pre-transfer editing ............................................................................................. 9
1.3.2 Post-transfer editing ................................................................. 10

1.4 Physiological rationale of editing ............................................ 15

1.4.1 Significance of ARSs editing function in vivo ....................... 15

1.4.2 Biological role of error-prone ARSs ...................................... 16

1.5 Purpose of this study ................................................................. 17

Chapter 2  Homologous trans-editing factors with broad substrate specificity prevent mistranslation caused by serine/threonine/homoserine misactivation ............... 18

2.1 Introduction ................................................................................ 18

2.2 Experimental Procedures ......................................................... 20

2.3 Results ...................................................................................... 23

2.3.1 Effect of E. coli proXp-y deletion on cell growth in the presence of elevated amino acid levels ..................................................... 23

2.3.2 In vitro deacylation by E. coli ProXp-y and B. parapertussis ProXp-z .......... 24

2.3.3 Deacylation activity of ProXp-y and ProXp-z towards cognate Thr-tRNA\textsubscript{Thr} ......................................................... 25

2.3.4 Deacylation activity of ProXp-y and ProXp-z towards aa-tRNA\textsuperscript{Ser} .......... 26

2.4 Discussion .................................................................................. 26

2.5 Acknowledgement ....................................................................... 31

Chapter 3: YbaK defends against oxidative stress in E. coli .......... 44

3.1 Introduction ................................................................................ 44
3.2 Materials and Methods ........................................................................................................ 46

3.3 Results and Discussion ..................................................................................................... 48

3.3.1 Expression levels of proS and ybak gene are regulated during stress. ........ 48

3.3.2 H₂O₂ induces growth defect in ybak null strain ................................................. 49

3.3.3 Cys amplifies the growth defect under ROS conditions ................................. 50

3.4 Conclusion .................................................................................................................... 53

Chapter 4: Mischarging and deacylation of non-protein amino acids in E. coli .......... 55

4.1 Introduction ..................................................................................................................... 55

4.2 Materials and Methods ............................................................................................... 57

4.3 Results and Discussion ............................................................................................... 58

4.3.1 BMAA is mischarged by SerRS, LysRS, AlaRS and ThrRS ........................... 58

4.3.2 BMAA-tRNA could be deacylated by ProXp-ST1 ........................................... 60

4.3.3 P5C is mischarged by ProRS ............................................................................... 61

4.3.4 EtSer- and iPrSer-tRNA could be deacylated by ProXp-ST1 ....................... 62

4.4 Conclusions .................................................................................................................. 64

Chapter 5: Ongoing projects and future direction ................................................................. 65

5.1 Probing the in vivo interacting partners of ProXp-ST1 and YbaK ....................... 65

5.2 Kinetic study of ARSs with non-cognate amino acids under stress condition ...... 67

5.3 X-ray crystallography studies of ProXp-ST1 ......................................................... 68
List of Tables

Table 1.1 Classes, editing substrates and domains of ARSs.................................................. 4
Table 1.2 Protein amino acids and closely related structural analogues that are misincorporated into proteins ........................................................................................................ 8
Table 2.1 Growth curve assays comparing E. coli WT and proXp-y null strains.............. 39
Table 2.2 Growth curve assays comparing E. coli WT and proXp-y null strains with ARS overexpressed.............................................................................................................................. 41
Table 4.1 Oxidative modification to amino acid residues found on proteins *in vivo*........57
Table 5.1 Buffer conditions from Wizard screen kit for ProXp-ST1 crystals............... 68
List of Figures

Figure 1.1 Two-step reaction catalyzed by tRNA synthetase........................................... 2
Figure 1.2 Pre- and post-transfer editing pathways ......................................................... 9
Figure 2.1 Domain structure of *E. coli* and *B. parapertussis* ProRS and the INS
superfamily..................................................................................................................32
Figure 2.2 Comparison of growth curves of *E. coli* WT and *proXp-y* null strains........33
Figure 2.3 Effect of Ser, Thr and Hse on the growth of *E. coli* WT and *proXp-y* null
strains with LysS, LysU or SerRS overexpression......................................................... 34
Figure 2.4 Deacylation of aa-tRNA by *E. coli* ProXp-y or *B. parapertussis* ProXp-z.....35
Figure 2.5 Deacylation of cognate Thr-tRNA$_{\text{Thr}}$ by *E. coli* ProXp-y or *B. parapertussis*
ProXp-z.............................................................................................................................36
Figure 2.6 Deacylation of Ser- and Hse-tRNA$_{\text{Ser}}$ by 0.5 μ *E. coli* ProXp-y or *B.
parapertussis* ProXp-z .................................................................................................... 37
Figure 2.7 Amino acids and tRNA specificity of ProXp-ST1 and ProXp-ST2................. 38
Figure 2.8 Effect of varying concentration of Thr, Ser and Hse on the growth of *E. coli*
WT and *proXp-y* null strain with LysS, LysU or SerRS overexpression is dose-dependent
........................................................................................................................................ 43
Figure 3.1 The expression level changes upon extracellular stress.............................49
Figure 3.2 Growth curve assays for *E. coli* WT and *ybak* null strains under different
concentrations of H$_2$O$_2$............................................................................................... 50
Figure 3.3 Growth curve assays for *E. coli* WT and *ybak* null strain under the indicated concentrations of H2O2 and Cys with or without overexpression of YbaK......................... 51

Figure 3.4 Intracellular Cys biochemistry ................................................................. 53

Figure 4.1 Structure of BMAA, Ser and Thr .............................................................. 59

Figure 4.2 Mischarging of BMAA by SerRS, LysS, LysU, ThrRS, AlaRS and ProRS.. 60

Figure 4.3 Deacylation of BMAA-tRNA by buffer or ProXp-ST1...............................61

Figure 4.4 Reactions of the proline metabolic pathway ........................................... 62

Figure 4.5 Mischarging of P5C-tRNAPro by 1 μM ProRS or 0.2 μM ProRS............... 62

Figure 4.6 Structures of Ser/Thr analogues linked with leaving groups in flexizyme reactions ................................................................................................................. 63

Figure 4.7 Deacylation of aa-tRNALys by ProXp-ST1............................................... 64

Figure 5.1 A schematic representation of CLIP procedure for identification of protein binding sites on RNAs in the context of intact cells........................................ 66

Figure 5.2 Template-switching activity of group II intron and retroviral RTs.......... 67

Figure 5.3 Growth curves of *E. coli* WT and *proXp-st1* null strain at 42 degree under Thr condition when LysU is overexpressed......................................................... 70
Chapter 1: Introduction

1.1 Fidelity of translation

Use of genetic information in all living cells comprises three major processes: replication, transcription and translation. Translation is a process of mRNA-guided protein synthesis using tRNA as adaptors. Firstly, a class of enzyme, called aminoacyl-tRNA synthetase (ARS), activates the cognate amino acid and attaches it onto the corresponding tRNA, forming aminoacyl-tRNA (aa-tRNA). Each ARS is specific for one amino acid. Secondly, the protein synthesis machinery uses aa-tRNAs as building blocks and delivers the correct aa-tRNA to the mRNA in the ribosome according to the genetic code to direct the growing polypeptide chain. Maintaining the fidelity of translation is critical for normal cell functions. The tolerance threshold of cellular protein synthesis is reported to be $10^{-5}$ to $10^{-3}$ in vivo (1), i.e., the error frequency is approximately $10^{-4}$ per codon. The inaccuracy of the aminoacylation step is cited to be greater than $10^{-4}$ and the ribosome decoding step is on the order of $10^{-4}$ in vitro. It is commonly believed that the fine-tuning of ARSs is one of the significant steps in protein quality control. Once mischarged by ARSs, aa-tRNAs are delivered by elongation factors (EF-Tu in bacteria and EF-1α in eukaryotes) with similar efficiency as the correct aa-tRNAs and misincorporated into nascent peptides (2,3).

1.2 Aminoacyl-tRNA synthetases

1.2.1 Aminoacylation reaction and ARSs class system
ARSs catalyze the aminoacylation reaction following a two-step mechanism (Fig. 1.1)

**Step 1: Activation**

\[
AA + ATP + aaRS \rightleftharpoons aaRS\cdot AA\cdot AMP + PP_i
\]

**Step 2: Transfer**

\[
tRNA \xrightarrow{aaRS\cdot AA\cdot AMP} AA\cdot tRNA + aaRS + AMP
\]

Figure 1.1 Two-step reaction catalyzed by tRNA synthetase.

In the first step, the amino acid attacks the Mg-ATP, forming the aminoacyl adenylate (aa-AMP) intermediate with the release of pyrophosphate. In the second step, either the 2’- or 3’-OH at the 3’-terminal A76 of tRNA attacks the enzyme bound adenylate, forming aa-AMP with the release of AMP. The binding order of ATP and amino acid are random, while four ARSs (GlnRS, GluRS, ArgRS and LysRS-I) are reported to be incapable of activating amino acid in the absence of tRNA (4).

ARSs are grouped into two classes based on the structural and functional characteristics (Table 1) (5). This classification focuses on the catalytic core domains of ARSs, but not other appended domains (6). Class I and class II ARSs contain 11 and 13 synthetases separately. Except for LysRS, each ARS belongs to either class I or class II, and the assignment is universal among all species.
ARSs catalyze the aminoacylation reaction following a two-step mechanism (Fig. 1.1)

**Step 1: Activation**

\[ AA + ATP + aaRS \rightleftharpoons aaRS\cdot AA\cdot AMP + PP_i \]

**Step 2: Transfer**

In the first step, the amino acid attacks the Mg-ATP, forming the aminoacyl adenylate (aa-AMP) intermediate with the release of pyrophosphate. In the second step, either the 2’- or 3’-OH at the 3’-terminal A76 of tRNA attacks the enzyme bound adenylate, forming aa-AMP with the release of AMP. The binding order of ATP and amino acid are random, while four ARSs (GlnRS, GluRS, ArgRS and LysRS-I) are reported to be incapable of activating amino acid in the absence of tRNA (4).

ARSs are grouped into two classes based on the structural and functional characteristics (Table 1) (5). This classification focuses on the catalytic core domains of ARSs, but not other appended domains (6). Class I and class II ARSs contain 11 and 13 synthetases separately. Except for LysRS, each ARS belongs to either class I or class II, and the assignment is universal among all species.
Class I synthetases contain a Rossmann fold motif in the catalytic domain, responsible for binding the acceptor stem of the tRNA (7,8). Two conserved sequences, “HIGH” and “KMSKS” are found in this domain and located near the α-phosphate of ATP. The Rossmann fold is separated by an inserted connective peptide (CP1) domain, which is enlarged in certain ARSs (IleRS, ValRS and LeuRS) as an edit site for mischarged tRNAs (9,10).
<table>
<thead>
<tr>
<th>Class</th>
<th>Edited amino acids</th>
<th>Editing domain</th>
<th>Mischarged amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cis</td>
<td>Trans</td>
</tr>
<tr>
<td>Class I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subclass Ia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CysRS</td>
<td>α</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IleRS</td>
<td>α, Hey, Cys, Val, Thr, Hse, α-aminobutyrate</td>
<td>CP1</td>
<td></td>
</tr>
<tr>
<td>LeuRS</td>
<td>α, Hey, Met, Ile, Val, norvaline, norleucine, Hse</td>
<td>CP1</td>
<td></td>
</tr>
<tr>
<td>MetRS</td>
<td>α, α2, Hey, active site</td>
<td></td>
<td>Norleucine, ethionine, S-nitroso-Hcy</td>
</tr>
<tr>
<td>ValRS</td>
<td>α, Thr, Cys, α-aminobutyrate, Ala, Hey</td>
<td>CP1</td>
<td>Ile, Thr, aminobutyrate</td>
</tr>
<tr>
<td>Subclass Ib</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ArgRS</td>
<td>α</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GlnRS</td>
<td>α</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GluRS</td>
<td>α</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LysRS</td>
<td>α</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subclass Ic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TrpRS</td>
<td>α2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TyrRS</td>
<td>α2</td>
<td>DTD</td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subclass IIa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GldRS</td>
<td>α2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HisRS</td>
<td>α2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProRS</td>
<td>α2, Ala, 4-hydroxyproline</td>
<td>INS</td>
<td>YbaK, ProXp-ala</td>
</tr>
<tr>
<td>SerRS</td>
<td>α2, Thr, Cys, serinehydroxamate</td>
<td>active site</td>
<td></td>
</tr>
<tr>
<td>ThrRS</td>
<td>α2, Ser</td>
<td>N2</td>
<td>ThrRS-ed</td>
</tr>
<tr>
<td>Subclass IIb</td>
<td></td>
<td></td>
<td>β-Hydroxynorvaline</td>
</tr>
<tr>
<td>AsnRS</td>
<td>α2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AspRS</td>
<td>α2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LysRS</td>
<td>α2, Orn, Hey, Hse</td>
<td>active site</td>
<td>Arg, Ala, Thr, Met, Leu, Cys, Ser</td>
</tr>
<tr>
<td>Subclass IIc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AlaRS</td>
<td>α, α4, Gld, Ser</td>
<td>C-terminus</td>
<td>AlaXp</td>
</tr>
<tr>
<td>GldRS</td>
<td>(αβ)2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PheRS</td>
<td>αβ3β4, Tyr, Ile</td>
<td>β3β4</td>
<td>Tyr</td>
</tr>
<tr>
<td>PylRS</td>
<td>α2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SepRS</td>
<td>α4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.1 Classes, editing substrates and domains of ARSs*

*Specific ARS is denoted by the three-letter amino acid designation.
Class II synthetases have a conserved catalytic domain consisting of a seven-stranded beta-sheet flanked by alpha-helices (11,12). Subunit organizations of these ARSs are different from class I: most are homodimers with several examples reported as monomeric or tetrameric (13-16). Three loosely conserved motifs are found in class II synthetases: motif 1 is found at the dimer interface while motif 2 and 3 responsible for substrate binding (17). HisRS, ProRS, SerRS and ThrRS are grouped as subclass IIA according to their similar catalytic core and C-terminal tRNA anticodon binding domain (18). Subclass IIB comprises AspRS, AsnRS and LysRS as a conserved oligonucleotide binding fold for anticodon binding is found in N-terminal extension (19-21). A central uracil base is conserved in the tRNAs of this subgroup. The class IIC synthetases are less conserved and contain only motif 2 and 3.

One class-specific distinction is that class I ARSs bind the tRNA acceptor helix on the minor groove site, while class II ARSs bind tRNA from the major groove. Another mechanistic difference is that all class I ARSs aminoacylate the 2’-OH of tRNA efficiently while class II enzymes activate the amino acids at the 3’-OH directly except for PheRS. In addition, the release of aa-tRNA is rate limiting for most class I synthetases, while in class II synthetases, the formation of aa-tRNA is the rate limiting step (22-24).

1.2.2 Specificity of ARSs

Proper pairing of the specific amino acid with corresponding tRNA by ARSs is the key to faithful translation. Specificity of ARSs is challenging as many structurally similar noncognate species exist for both tRNAs and amino acids.
The selection for tRNA-ARS binding is more specific as there is a large protein-tRNA interface present in the complex. The tRNAs are folded as an L-shape structures and recognized by ARSs by a fast and thermodynamic favored manner through the interactions with the anticodon or variable arm. However, the differential binding affinity is not enough for the selection of cognate tRNAs. Following the first stage of binding, the acceptor stem of tRNA interacts with the catalytic core of ARS during induce fit rearrangement (25-27). Thus the efficient aminoacylation reaction of cognate tRNA occurs more at the level of $k_{cat}$ than $K_M$. Identity elements of tRNA are necessary for both class I and class II synthetases (28). Most elements are found in the acceptor arm and the anticodon stem, e.g. N73 and N35 are two major discrimination elements (29). There are also other minor elements located throughout the tRNA that are specific to each ARS system. The acceptor stem with TψC arm and the anticodon stem with D arm in the L-shape tRNA are the two domains recognized by separate regions of ARS (30). The interactions between the active site of ARSs and the acceptor helix of tRNA are more specific in each synthetase subgroup, while the anticodon binding of tRNA-ARS can vary significantly among different systems (31,32).

Selection of the correct amino acid from the large pool of metabolites by ARS is more challenging, especially when similar amino acids are smaller than the cognate. Half of the canonical ARSs are not capable of differentiating the noncognate amino acids from the correct ones (Table 1). For example, *Escherichia coli* IleRS misactivates Val and hydrolyzes the adenylate in the presence of tRNA (33). Early studies also showed the hydrolytic activities of Thr by ValRS (34,35), Cys by IleRS (36,37), Gly and Ser by
AlaRS (38). A major challenge in amino acids selection includes homocysteine (Hcy), misactivated by four ARSs; Ser, misactivated by AlaRS, ThrRS and LysRS, and Thr, misactivated by IleRS, ValRS, LysRS and SerRS (39). Another notable case is that several ARSs misactivate larger amino acids than the cognate ones. AlaRS is shown to not only mischarge Gly onto tRNA^Ala, but also Ser (40). Similarly, PheRS mischarges Thr and ThrRS misactivates hydroxynorvaline (41-43). Additional hydrolytic domains are required in these cases.

Except for proteinogenic amino acids, a number of amino acid analogues are reported to be misactivated by many ARSs. Other than the Hcy mentioned above, norvaline is shown to be misactivated by LeuRS (44), homoserine by LysRS (45) and norleucine by MetRS and LeuRS (46,47) (Table 1.1). In addition, several amino acid analogues are found incorporated into proteins in vivo (Table 1.2). For example, canavanine, an analogue of Arg, is misincorporated into proteins and inhibits cell growth leading to apoptosis of cancer cells (48). Azetidine, found in Liliaceous, can be incorporated into proteins instead of Pro and impair the function of collagen. Oxidized amino acids are also shown to be substrate for ARSs. 3,4-dihydroxyphenylalanine (DOPA), o-hydroxyphenylalanine (o-tyrosine) and m-hydroxyphenylalanine (m-tyrosine) are the analogs result from hydroxyl radicals attack and are reported to be misincorporated into proteins. Moreover, ingestion and misincorporation of certain amino acid analogues, such as beta-N-methylamino-L-alanine (BMAA), can give rise to autoimmune disease symptoms. However, the relation between this analogue and ARSs remains unclear.
<table>
<thead>
<tr>
<th>Protein amino acids</th>
<th>Analogues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Canavanine</td>
</tr>
<tr>
<td>Glu</td>
<td>β-N-Methylaminoalanine (BMAA)</td>
</tr>
<tr>
<td>Leu</td>
<td>Leucine hydroxides</td>
</tr>
<tr>
<td>Phe</td>
<td>m-Tyrosine, o-tyrosine</td>
</tr>
<tr>
<td>Pro</td>
<td>Azetidine, proline hydroxides</td>
</tr>
<tr>
<td>Tyr</td>
<td>3,4-Dihydroxyphenylalanine (DOPA)</td>
</tr>
<tr>
<td>Val</td>
<td>Valine hydroxides</td>
</tr>
</tbody>
</table>

Table 1.2 Protein amino acids and closely related structural analogues that are misincorporated into proteins

1.3 Editing by aminoacyl-tRNA synthetases

To maintain high fidelity in translation, ARSs have evolved several editing pathways to clear the mismatched amino acid and tRNA. Ten ARSs are reported to have proofreading activities utilizing relatively variable mechanisms. Pre-transfer editing targets the misactivated aminoacyl-adenylate while post-transfer editing results in hydrolysis of the mischarged aa-tRNAs. Several pathways exist in editing process (Figure 1.2). ARSs may employ one or more pathways based on both structural and biochemical studies (49).
1.3.1 Pre-transfer editing

Once misactivation occurs, a “selective release” mechanism through nonenzymatic hydrolysis of aa-AMP in solution has been reported for pre-transfer editing. Alternatively, ARS may directly hydrolyze the aminoacyl adenylate via a tRNA-independent or tRNA-dependent catalytic activity.

Early pre-steady state kinetic studies on class I IleRS first elucidated the enzyme-catalyzed tRNA-independent hydrolysis of aminoacyl-adenylates, as the half-life of Val-AMP in solution is much longer than the rate of destruction measured in these experiments (36). Thus, a mechanism other than selective release is confirmed for IleRS. Class I ValRS, LeuRS and class II ProRS are also shown later to employ this mechanism (23,50,51). TRNA-independent pre-transfer editing is also used by ThrRS, SerRS, ProRS, MetRS, and class II LysRS, though distinct cyclization pathways are employed by latter two enzymes for misactivated Hcy and homoserine (Hse)/ornithine (Orn) separately (39,52).
Compared to LeuRS and ValRS, tRNA-dependent pre-transfer editing contributes significantly in IleRS, as the $k_{cat}$ for tRNA-dependent pre-transfer editing is ten-fold higher than for tRNA-independent editing (23). A translocation model has been proposed for IleRS wherein Val-AMP is shuttled to the CP1 editing site during pre-transfer editing (53). Several structural studies provide evidence supporting this model as the CP1 domain is rotated when tRNA binds and 5’-o-[N-(L-norvalyl)sulfamoyladenylate]adenosine (NvaAMS) binds both the aminoacylation and the editing site (54,55). In contrast, as MetRS, SerRS and class II LysRS lack the post-transfer editing domain, the pre-transfer editing is restricted to the aminoacylation site in these synthetases.

1.3.2 Post-transfer editing

Prior to studies reporting hydrolysis of Val-tRNA$^{Ile}$ by IleRS and Ile-tRNA$^{Phe}$ by PheRS, a classic “double-sieve” model was proposed by Fersht in 1977 based on the study of Ile-tRNA$^{Val}$ (56). Two levels of discrimination exist in ARSs: noncognate amino acids are first excluded by the aminoacylation site (coarse sieve) based on their size, and a second editing site serves as another sieve to deacylate the smaller noncognate substrates. This model has been validated in several ARS systems including LeuRS and class II ProRS, AlaRS and ThrRS (9). In addition to the in cis editing domain of ARSs, freestanding post-transfer domains that hydrolyze mischarged tRNAs in trans have been discovered throughout genome (57).

1.3.2.1 Cis-editing domains
A common CP1 domain that possesses the post-transfer editing function was found in three class I synthetases that have post-transfer editing activities. Despite the domain structures in various synthetase systems, the spatial distance between synthetic and editing sites is approximately 30-40 Å (54,58-60). How the aminoacyl-tRNAs reach the editing sites is controversial. Two mechanisms have been proposed for post-transfer editing. A translocation model suggests that the 3’ end of tRNA shuttles between catalytic and editing sites as a result of a conformational change of the ARS-tRNA complex (50,55). This model is supported by structural studies of IleRS in complex with tRNA$^{\text{Ile}}$ where the CP1 domain is found to rotate upon tRNA binding (54), and in LeuRS with tRNA$^{\text{Leu}}$ where the 3’ acceptor of tRNA is trapped in the CP1 domain (6,55). Alternatively, another model has been reported wherein mischarged tRNAs first dissociate from the synthetase and then rebind in the editing domain for editing. This model is provided in PheRS and has been shown to be significant in the post-transfer editing of Thr-tRNA$^{\text{Phe}}$ (61). However, although the dissociation and rebinding is a possible mechanism for the CP1 domain, it may not contribute significantly to the wild-type synthetases. It is likely that due to the rate-limiting product release step in class I synthetases, the translocation model plays the primary role in post-transfer editing by these enzymes.

In contrast to class I synthetases, post-transfer editing domains of class II PheRS, AlaRS, ThrRS and ProRS are much less conserved. *Escherichia coli* PheRS possesses the editing function in the beta subunit which is 40 Å away from the synthetic site (41,62,63). However, various forms of PheRS are found in different species and cellular
compartments (64). Many archaeal and eukaryal PheRSs share less homology of the critical residues comparing to bacterial ones (65), and mitochondrial PheRSs lack the post-transfer editing subunit (66).

Post-transfer editing domains in ThrRS and AlaRS share homology. The N2 domain of bacterial ThrRS is capable of deacylating Ser-tRNA\textsuperscript{Thr}, which is often absent in mitochondrial and archaeal ThrRSs (67,68). Crystal structures of bacterial ThrRS bound to post-transfer analog elucidate the general base catalysis mechanism (42,69). Two water molecules exist in the editing site and facilitate the conserved His and Lys residues in nucleophilic attack and transition-state stabilization. When Thr is bound instead of Ser, one molecule is excluded in the editing site. Similar mechanisms are proposed for ProRS, PheRS and LeuRS based on computational studies (70-72).

AlaRS mischarges Ser and Gly onto tRNA\textsuperscript{Ala} and an appended editing domain is responsible to clear both of these species (73). As suggested by the unique structure of AlaRS, the flexible movement of the acceptor arm between the synthetic and editing sites is not straightforward and the translocation model is thus not likely the primary pathway (74,75). A unique G3:U70 base pair of tRNA\textsuperscript{Ala} serves as the primary identity element and determines the specificity of Ser-tRNA\textsuperscript{Ala} recognition (76). The lack of anticodon binding of AlaRS also supports the dissociation and rebinding model.

In ProRS, the aminoacylation site mischarges Ala and Cys onto tRNA\textsuperscript{Pro}, while an inserted domain (INS), which is present in most bacterial ProRS, is capable of deacylating Ala-tRNA\textsuperscript{Pro} in cis (77,78). Structural studies reveal that the synthetic and editing sites in ProRS are linked by a flexible peptide and are separated from each other.
by approximately 35 Å. The docking model is consistent with the translocation mechanism, wherein the acceptor end flips between two catalytic site of ProRS without interfering with tRNA binding. Besides, ProRS with a full-length INS, another isoform of ProRS containing a mini-INS is also found in bacteria. This mini-INS is not active toward mischarged tRNAs (57,79).

1.3.2.2 Trans-editing factors

Freestanding editing proteins are homologs of the editing domains of ARSs found throughout three domains of life and hydrolyze the mischarged tRNAs after their dissociation (57). As product release is rate-limiting for class I synthetases, all trans-editing factors are found in class II ARS systems. One exception is the D-Tyr-tRNA\textsuperscript{Tyr} deacylase (DTD), which hydrolyzes D-Tyr-tRNA\textsuperscript{Tyr} misacylated by TyrRS (80,81). However, DTD shares sequence and structural homology with the editing domain of archaeal ThrRS, which has editing specificity for Ser-tRNA\textsuperscript{Thr} (261-263). DTDs are widely distributed in all kingdoms and are essential for cell viability (69). The homolog of the editing domain of ThrRS is also found as a freestanding protein ThrRS-ed in achaea (68).

AlaX, a freestanding homolog of the editing domain of AlaRS, is capable of deacylating mischarged Ser- and Gly-tRNA\textsuperscript{Ala} formed by AlaRS, and is conserved in all three domains of life (57). AlaXs are classified into three isoforms according to their length: AlaX-S, AlaX-M and AlaX-L (82). AlaXs in some species have deacylation ability only towards Ser-tRNA\textsuperscript{Ala}, and lack the capability to hydrolyze of Gly-tRNA\textsuperscript{Ala} (83), suggesting Ser-tRNA\textsuperscript{Ala} is the primary target for these freestanding proteins.
Recently, a study has shown that the more ancient archaeal AlaX-S has a broad activity towards Ser-tRNA$^\text{Ala}$, Ser-tRNA$^\text{Pro}$, and Ser-tRNA$^\text{Thr}$, but not Gly-tRNA$^\text{Ala}$. Considering the phylogenetic distribution, it is likely that AlaX-S is the functional ancestor of AlaRS and ThrRS editing domain (Novoa E. et al., unpublished). Since the aminoacylation site of AlaRS could not discriminate the larger Ser from the cognate Ala, and the substitution of Ala by Ser is extremely deleterious, AlaXSs are maintained as a result of strong evolutionary pressure even in higher organisms (40).

ProRS has the most diverse post-transfer mechanisms. A family of freestanding homologs of the bacterial INS including YbaK, ProXp-ala (previously known as PrdX), ProXp-x (previous named ProX), ProXp-y (previous YeaK) and ProXp-z (previous PA2301) are found across all three kingdoms and are collectively known as the YbaK superfamily (85). YbaK has been shown to deacylate Cys-tRNA$^\text{Pro}$ formed by the aminoacylation site of ProRS both in vitro and in vivo (86-88). A ternary complex forms between YbaK, ProRS and Cys-tRNA$^\text{Pro}$ as revealed by a crosslinking study (89), which ensures that the activity of YbaK is dedicated to Cys-tRNA$^\text{Pro}$. Deacylation of YbaK occurs through a distinct cyclization pathway involving the thiol chemistry of Cys (90,91).

The YbaK trans-editing factor, together with the aminoacylation active site and cis-editing domain, constitute a “triple sieve” mechanism to ensure fidelity in Pro codon translation. The complexity of this system reflects the high requirement of living cells to maintain faithful genetic information flow. ProXp-ala has the same substrate specificity as the INS domain. In some organisms that encode ProRS without post-transfer editing
function, including *Clostridium sticklandii*, ProXp-ala substitutes for the INS domain to clear mischarged Ala-tRNA$_{\text{Pro}}$ in trans (90,92). In organisms that also encode YbaK, such as *Caulobacter crescentus*, two trans-editing domains and the synthetic active site of ProRS, constitute an alternative triple sieve (85). In addition, ProXp-ala is found in higher organisms including *homo sapiens* (93). Additional free-standing ProRS trans-editing factors ProXp-x, ProXp-y and ProXp-z are phylogenetically related to the INS domain, but the substrate specificity has not been reported to date.

1.4 Physiological rationale of editing

The high fidelity of aminoacylation in protein synthesis is ensured by various editing mechanisms of ARSs and their homologs, providing the maximal survival advantage for living cells. The importance of editing is shown by multiple studies in different organisms from bacteria to human (94). Understanding the extent of tolerance of mistranslation under different local conditions, in different cells or cellular compartments could be of special significance to certain disease, such as cancer and Alzheimer’s disease.

1.4.1 Significance of ARSs editing function in vivo

In most cases, mutations in ARS editing domains result in adverse effects on cell growth, especially where stress is added to the living environment. *E. coli* ValRS with impaired proofreading ability results in a severe growth defect in the presence of elevated levels of $\alpha$-aminobutyrate, a non-protein amino acid misactivated by ValRS (95). Editing
defective IleRS gives rise to a wide-range of consequences in *E. coli* and leads to higher DNA mutation rates (96). Mistranslation caused by editing defective AlaRS in mammalian cells results in accumulation of misfolded proteins and subsequent severe neurodegenerative phenotype (97). In addition to AlaRS, an editing deficient mutant of ValRS triggers an apoptosis-like response and changes cellular morphology in mouse cells (97).

However, cells can survive different levels of misincorporation into proteins and the capability to tolerant translational inaccuracy varies between cells and compartments. 10% substitution of Asn for Asp is found in actively growing *E. coli* with editing deficient AspRS expressed (2). The A734E mutant of AlaRS results ingreatly decreased translational fidelity and significant damage in nondeniding cerebellar Purkinje cells in the brain, but is not shown harmful to all other cells tested (97). Another example is the editing deficient PheRS. While substitution wild-type PheRS with the mutant synthetase in *E. coli* and *S. cerevisiae* has little effect on growing in rich media, the replacement in yeast mitochondria is detrimental (64). Thus, the requirement of editing highly depends on the biological context.

1.4.2 Biological role of error-prone ARSs

The levels of mistranslation by ARSs are shown to be altered under particular conditions. The editing ability of *E. coli* ThrRS is impaired under oxidative stress where the Cys residue in the active site is oxidized (98). Increasing amounts of mischarged Ser-tRNA_{Thr} arises in this circumstance and results in misfolded proteins caused by Ser misincorporation. ThrRS editing function could be restored once the stress is removed;
thus, it is possible that the increased Ser is involved in a Cys signaling pathway by providing more sites for phosphorylation. It is more common for ARSs to change their tRNA specificity depending on the local environment. The imbalance of ARS and tRNA would lead to mischarging of noncognate tRNA \textit{in vivo}. The mischarging of noncognate tRNA caused by overexpression of GlnRS is suppressed with co-expression of tRNA$^{Gln}$\(^{(99)}\). LeuRS in \textit{Candida albicans} mischarges tRNA$^{Ser}$ to a level of 3% to 5% \textit{in vivo}, leading to significant phenotype diversity due to ambiguous decoding \((100)\). More recent work on human HeLa cells shows that Met misacylation is elevated 10 fold under ROS conditions compared to normal growth conditions \((101)\). The misacylation is likely caused by MetRS as both \textit{E. coli} and \textit{S. cerevisiae} MetRS are found to be capable of misacylating different tRNA species in a tunable manner \((102,103)\). Therefore, the fidelity of aminoacylation could be adaptive under various conditions, particularly stress, and be actively regulated by cells to suit the environment.

1.5 Purpose of this study

This thesis describes a screen used to establish the substrate specificity of \textit{trans}-editing factor ProXp-y and biochemical studies on freestanding proteins (YbaK, ProXp-y and ProXp-z) of the INS superfamily. Various \textit{in vivo} assays were performed to investigate the possible interactions of \textit{trans}-editing domains with different ARSs under normal and stress conditions. In addition, several non-natural amino acid analogues were tested as the substrates for specific ARS and ProXp-y. Ongoing studies include exploring the tRNA and ARS binding partners of YbaK and ProXp-y in cells.
Chapter 2  Homologous trans-editing factors with broad substrate specificity prevent mistranslation caused by serine/threonine/homoserine misactivation

2.1 Introduction

A high level of accuracy in protein synthesis is essential for normal cell function and proliferation. A critical step in this process is pairing the correct amino acid from the pool of 20 natural amino acids to the cognate tRNA species by aminoacyl-tRNA synthetases (ARSs). ARSs catalyze the aminoacyl-tRNA formation (aa-tRNA) in two steps involving amino acid activation and transfer of the activated amino acid to tRNAs. Although ARSs have evolved to exhibit specific tRNA recognition capabilities with an estimated error frequency of $10^{-6}$, a greater number of mistakes arise from the lack of discrimination of near-cognate amino acids, with an estimated error rate of $10^{-4}$ to $10^{-5}$ at this step (94). Misincorporation of amino acids into proteins can be harmful to both eukaryotic and prokaryotic cells (4,40). For example, a mutation in the editing site of alanyl-tRNA synthetase (AlaRS), which results in only a 2-fold increase in misacylation in vitro, results in a severe neurodegeneration phenotype in mice (97). Cellular degradation and apoptosis caused by a mutation in the editing domain of valyl-tRNA synthetase (ValRS) has been reported in human cells (104). In addition, editing defects in bacteria often results in slower growth rates, delayed cell growth, or even cell death (46,96,105-107).
Quality control of aa-tRNA formation is achieved by hydrolysis of the aminoacyl-adenylate (“pre-transfer editing”) or deacylation of the mischarged aa-tRNAs (“post-transfer editing”) (39, 49). Editing mechanisms are employed by both classes of ARSs, and seven out of twenty-two ARSs possess post-transfer editing sites that are distinct from the aminoacylation active site. The connetive polypeptide 1 (CP1) editing domain is found in class I isoleucyl-tRNA synthetase (IleRS), leucyl-tRNA synthetase (LeuRS), and ValRS (108-112). Editing domains of class II ARSs are more diverse and include the N-terminal domain (N2) of threonyl-tRNA synthetase (ThrRS) (113), the related editing domain of AlaRS (114), the β3/β4 domain of phenylalanine-tRNA synthetase (115), and the insertion domain (INS) of prolyl-tRNA synthetase (ProRS) (77, 79).

In addition to the cis-editing domain appended to ARSs, freestanding homologs of editing domains are distributed throughout organisms in all three kingdoms of life as additional checkpoints to maintain translational fidelity in trans. Autonomous trans-editing factors that are evolutionary related to class II AlaRS, ThrRS and ProRS have been identified. A homolog of the AlaRS editing domain, AlaXp, is widely distributed and is shown to hydrolyze Ser-tRNA_{Ala} (57, 116). ThrRS-ed is an autonomous editing domain of Ser-tRNA_{Thr} in crenarchaeal genomes, where thrS genes are truncated (68). Bacterial ProRS INS domain homologs include five proteins previously named YbaK, ProXp-ala, ProXp-x, ProXp-y (annotated YeaK) and ProXp-z (annotated PA2301) (85) (Fig. 2.1). These domains together with the INS domain of ProRS are collectively known as the “INS superfamily”. While the INS domain edits mischarged Ala-tRNA^{Pro}, YbaK deacylates Cys-tRNA^{Pro}, which is formed in the active site of ProRS due to the similar
size of Cys and Pro (87). ProXp-ala is capable of clearing Ala-tRNA<sub>Pro</sub> and compensates for the lack of an INS post-transfer editing domain in many bacteria (57). In contrast, functions for ProXp-y and ProXp-z have not yet been reported. In this study, we explore the in vivo and in vitro substrate specificities of ProXp-y and ProXp-z for the first time.

2.2 Experimental Procedures

_Materials and strains_—All amino acids and chemicals were purchased from Sigma-Aldrich unless otherwise noted. [α<sup>32</sup>P]ATP was from PerkinElmer Life Sciences. *Escherichia coli* strain BW25113 (lac<sup>il</sup> rrnB<sub>T14</sub> ΔlacZ<sub>W16</sub> hsdR514 ΔaraBAD<sub>AH33</sub> ΔrhaBAD<sub>LD78</sub>) and the *proxp-y* null strain (Δyeak, JW1776) from the Keio collection were obtained from the Coli Genetic Stock Center (117). The plasmid pCP20, which carries the FLP recombinase, was used to remove the kanamycin cassette from JW1776 mutants (118). Plasmids encoding each ARS gene in vector pCA24N (-gfp) were obtained from the ASKA collection (119).

_Growth curve assays_—M9 minimal media supplemented with 0.4% glucose and indicated concentrations of amino acids was used in all growth curve assays. *E. coli* cells were grown in LB media overnight, collected and washed twice with minimal media. For strains carrying ARS expression plasmids, 20 μg/ml chloramphenicol was added to rich media. Cells were diluted to an O.D. 600 of 0.01, and grown at 37 °C in M9 minimal media in the presence of Biolog Redox Dye Mix MA, 0.005 mM β-D-1-thiogalactopyranoside (IPTG) and in the absence and presence of the indicate
concentration of specific amino acids. Cell growth was automated monitored at 37 °C for
48 h using the Omnilog® system (Biolog Inc., Hayward, CA).

**Enzyme preparation**—The *E. coli* proxp-y gene was cloned into pET15b with a N-
terminal His-tag and overexpressed in *E. coli* BL21 (DE3) by addition of 0.1 mM IPTG
and incubation for 12 h at room temperature. The *Bordetella parapertussis* pa2301 gene
sequence encoding ProXp-z was codon-optimized for expression in *E. coli* and cloned
into pET15b (Novagen) by Genewiz. ProXp-z overexpression was carried out in *E. coli*
BL21 (DE3) as described for ProXp-y. To prepare ProXp-y and ProXp-z proteins, cells
were pelleted by centrifugation at 6000 rpm for 20 min at 4 °C. Buffer containing 50 mM
NaH₂PO₄ (pH 7), 300 mM NaCl, and 10 mM β-mercaptoethanol was used for cell lysis,
followed by sonication. The His-tagged proteins were purified using HIS-select® nickel
resin, and eluted with an imidazole gradient. Peak fractions containing protein were
concentrated using Amicon ultra centrifugal filters and stored in 50 mM NaH₂PO₄ (pH 7),
150 mM NaCl, and 40% glycerol at -20 °C. Wild type (WT) *E. coli* ThrRS, *E. coli*
C666A/Q584H AlaRS (114), *E. coli* K279A ProRS (120), *E. coli* T222P ValRS (121), *E.
coli* lysyl-tRNA synthetase (LysRS), and *E. coli* tRNA nucleotidyltransferase were
prepared as previously described (122). Protein concentrations were determined by the
Bradford assay (Bio-Rad).

**Substrate preparation**—*E. coli* tRNA₅er, tRNA₅ala, tRNA₅pro, tRNA₅ile, tRNA₅val,
tRNA₅thr and tRNA₅lys were prepared by in vitro transcription using T7 RNA polymerase
as previously described (77). All tRNAs were 3’-[³²P]-labeled by tRNA
nucleotidyltransferase (122). Preparation of *E. coli* Ser-tRNA₅er, Ser-tRNA₅thr, Ser-
tRNA\textsuperscript{Pro}, and Ile-tRNA\textsuperscript{Ile} was carried out using biotinylated dinitro-flexizyme (dFx) and Ser-3, 5'-dinitrobenzyl ester (Ser-DBE) or Ile-DBE as described. Thr-tRNA\textsuperscript{Ser} and Thr-tRNA\textsuperscript{Ile} were prepared using enhanced-flexizyme (eFx) and Thr-cyanomethyl ester (Thr-CBT) (123). \textit{E. coli} C666A/Q584H AlaRS was used for preparation of \textit{E. coli} Ala-, Gly-, and Ser-tRNA\textsuperscript{Ala}, \textit{E. coli} K279A ProRS for preparation of Ala- and Pro-tRNA\textsuperscript{Pro}, \textit{E. coli} T222P ValRS for preparation of Val- and Thr-tRNA\textsuperscript{Val}, \textit{E. coli} LysRS for preparation of Lys-, Ser-, and Thr- tRNA\textsuperscript{Lys}, \textit{E. coli} ThrRS for preparation of Thr-tRNA\textsuperscript{Thr} and \textit{E. coli} SerRS for preparation of homoserine (Hse)-tRNA\textsuperscript{Ser}. Following aminoacylation reactions using standard protocols (85), aa-tRNAs were phenol-chloroform-extracted followed by ethanol precipitation. Substrates for deacylation assays were stored as pellets at -80 °C.

\textit{Deacylation Assays—In vitro} deacylation assays were performed at 20-25 °C in reactions containing ~0.75 μM \textsuperscript{32}P-labeled aa-tRNA, 50 mM HEPES (pH 7.5), 2 mM dithiothreitol, 20 mM KCl, 5 mM MgCl\textsubscript{2}, 0.1 mg/ml bovine serum albumin and 15 μg/ml inorganic pyrophosphatase. Reactions were initiated by addition of 0.5 μM \textit{E. coli} ProXp-\textit{y} or \textit{B. parapertussis} ProXp-\textit{z}, and quenched at indicated time points by putting 2 μl of the reaction into 6 μl of a solution containing 200 mM NaOAc (pH 5.5) and 0.5 U/μl of P1 nuclease. The reaction products (aa-[\textsuperscript{32}P]AMP and [\textsuperscript{32}P]AMP) were analyzed on polyethyleneimine-cellulose TLC plates as previously described (122). Graphical analysis was performed using SigmaPlot (Systat Software, San Jose, CA) with error bars representing the standard deviation (S.D.) of triplicate measurements. Reactions to measure spontaneous (no enzyme) hydrolysis were performed for each aa-tRNA substrate and subtracted from deacylation reactions with enzyme.
2.3 Results

2.3.1 Effect of *E. coli* proXp-y deletion on cell growth in the presence of elevated amino acid levels

As members of the ProRS INS superfamily, we first tested the *in vitro* editing activities of *E. coli* ProXp-y and *B. parapertussis* ProXp-z using Pro-, Ala- and Cys-tRNA\(^{\text{Pro}}\) substrates. However, no deacylation activity was detected *in vitro* (data not shown). We took advantage of the availability of an *E. coli* proxp-y null strain (Keio collection), which was viable under normal growth conditions, to perform an *in vivo* screen in the presence of elevated levels of each of each of the twenty genetically encoded amino acids. We observed a significant and reproducible increase in the lag phase of the null strain relative to the WT strain only in the presence of elevated levels of Ser and Leu (Table S1). Based on these results we also tested structurally related homoserine (Hse) and observed a similar phenotype (Table S1).

We hypothesized that increased expression of certain error prone synthetases together with elevated levels of a near cognate amino acid may amplify any growth defect due to mistranslation in the null strain. The WT and proxp-y null strains were transformed with overexpression plasmids encoding each of the *E. coli* class II ARS genes. The *E. coli* genome encodes two class II LysRS genes, constitutive *lysS* and stress inducible *lysU*. The two corresponding enzymes share 88% identity, but possess different affinities for Lys and different thermostabilities (124). The mischarging activities of these two LysRS isoforms have not been investigated extensively and both were included in
our screen. The growth of these strains was measured in the presence of elevated levels of each of the amino acids tested in the first-screen. Among the 231 growth conditions screened, growth defects were observed in the null strain in the presence of Ser when AlaRS, LysS or LysU were overexpressed and with Thr when LysU was overexpressed (Fig 2.2, Table 2.2). Adding Hse also results in a phenotype when LysS, LysU and SerRS were overexpressed (Table 2.2).

AlaRS is known to misactivate Ser (38), and LysRS has been reported to misactivate Ser, Thr and Hse (45). Although *E. coli* SerRS has been reported not to misactivate Hse, in our hands, mischarging of tRNA<sup>Ser</sup> is observed in the presence of 5 mM Hse (data not shown). The differences observed between the WT and *proxp-y* null strains under the conditions shown in Fig. 2.2 and Table 2.1 suggest the potential roles of ProXp-y in editing of tRNAs mischarged by LysRS, AlaRS, SerRS and possibly ThrRS. LysU overexpression appeared to result in a much larger growth defect than LysS overexpression. To further explore these differences, a dose-dependent analysis was carried out by varying the concentrations of added amino acids (Fig 2.1). As summarized in Fig. 2.3 for the 5 mM concentrations of Ser, Thr and Hse, the *proxp-y* null strain expressing the stress-inducible *lysU* gene showed a dramatic increase in growth time to half-maximal O.D. 600 compared with the WT strain. The largest effect was observed for Hse, where no growth was observed with LysU in the null strain. SerRS overexpression in *proxp-y* null strain showed a similar growth defect in the presence of Hse. Taken together, these results indicate that LysU may be more error-prone than LysS, and ProXp-y may be required under physiological conditions where the *lysU* gene is induced
2.3.2 *In vitro* deacylation by *E. coli* ProXp-y and *B. parapertussis* ProXp-z

Based on the *in vivo* results using the proxp-y null strain, we next examined the editing capability of both *E. coli* ProXp-y and the closely related ProXp-z *in vitro*. As the later domain is not encoded in *E. coli*, we cloned and purified the homolog from *B. parapertussis* for these studies. For both proteins, moderate levels of deacylation were observed for mischarged Ser-tRNA^Lys^, Ser-tRNA^Ala^ and Hse-tRNA^Lys^ with more robust activities measured for Thr-tRNA^Lys^ (Fig. 2.4A and 2.4B). No deacylation of cognate Lys-tRNA^Lys^ and Ala-tRNA^Ala^ was observed. In addition to formation of mischarged Ser-tRNA^Ala^, Ser-tRNA^Lys^, and Thr-tRNA^Lys^ by AlaRS and LysRS, Thr and Ser are known to be misactivated by ValRS, IleRS and ThrRS to form Thr-tRNA^Val^, Thr-tRNA^Ile^ and Ser-tRNA^Thr^. Thus, these substrates were prepared to test whether the activities of ProXp-y and ProXp-z extend beyond tRNA^Ala^ and tRNA^Lys^. Deacylation assays showed that ProXp-y and ProXp-z also edit Thr-tRNA^Val^, Thr-tRNA^Ile^ and Ser-tRNA^Thr^, whereas cognate Val-tRNA^Val^ and Ile-tRNA^Ile^ species are not deacylated (Fig. 2.4C and 2.4D). These results suggest the role of ProXp-y and ProXp-z as general deacylases with dual specificities for mischarged Ser/Thr-tRNAs.

2.3.3 Deacylation activity of ProXp-y and ProXp-z towards cognate Thr-tRNA^Thr^

Based on the robust deacylation of various Thr-tRNAs by ProXp-y and ProXp-z, we wondered whether cognate Thr-tRNA^Thr^ was also hydrolyzed by these two enzymes. Indeed, Thr-tRNA^Thr^ is effectively deacylated by ProXp-y and ProXp-z (Fig. 2.5). However, the presence of even sub-stoichiometric amount of ThrRS protects Thr-
tRNA\textsuperscript{Thr} from hydrolysis, whereas excess ProRS does not. This result suggests that Thr-tRNA\textsuperscript{Thr} is not deacylated by these two free-standing editing domains in a physiological context.

2.3.4 Deacylation activity of ProXp-y and ProXp-z towards aa-tRNA\textsuperscript{Ser}

In contrast to deacylation of Thr-tRNA\textsuperscript{Thr} by ProXp-y and ProXp-z, Ser-tRNA\textsuperscript{Ser} is not efficiently cleared by ProXp-z, although it is a moderate substrate for ProXp-y (Fig. 2.6). Compared with deacylation of Ser-tRNA\textsuperscript{Lys}, Ser-tRNA\textsuperscript{Ser} is deacylated with a 3-fold and 50-fold reduced rate by ProXp-y and ProXp-z respectively. In both cases, despite the cell-based results showing a dramatic growth defect with added Hse (Fig. 2.3), very similar deacylation rates were observed for Hse-tRNA\textsuperscript{Ser} relative to Ser-tRNA\textsuperscript{Ser}. In addition, Thr-tRNA\textsuperscript{Ser} is a good substrate for ProXp-y, but not ProXp-z. This is the only major difference in tRNA specificity observed for these two homologs.

2.4 Discussion

A key step in maintenance of high fidelity in protein synthesis, catalysis of aa-tRNA formation by ARSs is balanced between efficiency and specificity. While the active sites of ARSs reject most noncognate amino acids in the “primary sieve”, their ability to differentiate protein-encoding and non-genetically encoded amino acids from cognate substrates is challenged by smaller and structurally related amino acids. Thus, “double-“ and “triple-sieve” editing mechanisms have evolved in all three kingdoms of life and become essential under certain growth conditions, including environmental stress.
Previously, two *trans*-editing domain homologs of the bacterial ProRS INS domain, YbaK and ProXP-Ala, were shown to deacylate known ProRS editing errors, with specificity for Cys- and Ala-tRNA<sub>Pro</sub>, respectively. Whereas YbaK gains tRNA<sub>Pro</sub> specificity through interaction with ProRS (89), ProXP-Ala recognizes tRNA<sub>Pro</sub> through interaction with the same acceptor stem identity elements as bacterial ProRS (85). Here, we show that two other INS domain homologs, ProXP-x and ProXP-y, have specificity for Ser/Hse- and Thr-tRNAs and are therefore unlikely to clear ProRS mischarging errors. Instead, based on in vitro deacylation studies, these enzymes are capable of clearing mischarging errors caused by LysRS (Ser/Hse- and Thr-tRNA<sub>Lys</sub>), AlaRS (Ser-tRNA<sub>Ala</sub>), and ThrRS (Ser-tRNA<sub>Thr</sub>), as well as class I IleRS (Thr-tRNA<sub>Ile</sub>) and ValRS (Thr-tRNA<sub>Val</sub>). Given their in vitro specificity, we renamed ProXP-x and ProXP-y as ProXP-ST1 and ProXP-ST2 (Fig. 2.7).

With the exception of LysRS and SerRS, the other synthetases listed in Fig. 2.7 all have known *cis*-editing capability and in some cases *trans*-editing domains also exist. For example, redundant mechanisms exist for correcting errors made by AlaRS due to the similar structure of Ser relative to cognate Ala (40). The editing domain of AlaRS, as well as the *trans*-editing domain AlaXP display similar capabilities to deacylate Ser-tRNA<sub>Ala</sub> (57). Since ProXP-ST1 and –ST2 also deacylate Ser-tRNA<sub>Ala</sub>, we investigated the distribution of the three *trans*-editing proteins. Of 1475 genomes examined representing all kingdoms of life, 67% encode at least one of these *trans*-editing domains; this number is 57% when only Bacteria are considered. Since ProXP-ST1 and –ST2 are not encoded in Archaea and are only present in very few eukaryotes, only Bacteria were
considered further. No Bacteria encode all three trans-editing domains and there is very little overlap between ProXp-ST1 and ProXp-ST2 (only 1 species encodes both) and between ProXp-ST2 and AlaX (7 species encode both). However, 110 species (~11%) encode both AlaX and ProXp-ST1. Taken together, these data suggest that Ser-tRNA\textsuperscript{Ala} may not be a major substrate for the ProXp family of editing domains in vivo.

The ability to differentiate Ser and Thr from each other and from other amino acids appears to be particularly challenging for ARSs, as at least 6 of the 20 synthetases misactivate these substrates. ThrRS readily misactivates Ser and in most species, ThrRS encodes an N-terminal post-transfer editing domain to clear Ser-tRNA\textsuperscript{Thr}. In several crenarchaeal species, the editing domain ThrRS-ed is encoded as separate trans-editing factor (68). Interestingly, \textit{E. coli} ThrRS editing function is severely impaired under oxidative stress conditions and this phenotype is attributed to an oxidized Cys residue in the editing domain active site (98). We hypothesize that ProXp-ST1 may be required to clear Ser-tRNA\textsuperscript{Thr} under conditions where ThrRS editing function is lost.

All trans-editing factors identified to date are related to class II synthetase editing domains (AlaRS, ThrRS, and ProRS) and are believed to assist in editing of tRNAs mischarged by these same synthetases. However, here we show that ProXp-type editing domains are able to deacylate a wide variety of tRNAs including those potentially mischarged by class I IleRS and ValRS. Although Thr-tRNA\textsuperscript{Ile} and –tRNA\textsuperscript{Val} were robustly deacylated by both ProXp-ST1 and –ST2, both of these synthetases possess editing domains that clear these mischarged tRNAs under normal conditions. Although it is possible that the ProXp-type domains perform a redundant editing function under
certain conditions, differences in aminoacylation kinetics between class I and class II ARSs suggest that trans-editing may not be required for tRNAs charged by class I synthetases (9). Product release is rapid in class II ARSs, unlike class I enzymes where this step is rate-limiting (22,61). Thus, class I ARSs remain bound to their substrates long enough to edit any misacylated tRNA prior to release and would not normally require proofreading by a trans-editing factor.

On the other hand, LysRS is a class II synthetase with no known post-transfer editing capability, despite reported mischarging of a variety of protein and non-protein amino acids in vitro (45). Based on the results reported here, we propose that mischarging errors caused by LysRS isoform LysU are the primary substrates for ProXp-ST1 and –ST2 (Fig. 2.7). We hypothesize that these domains are required under conditions of cellular stress, which is when LysU levels are elevated. We have shown that when cellular Leu levels are elevated, the ProXp-ST1 null strain displays a reduced growth phenotype even though ProXp-ST1 doesn’t edit Leu-tRNAs (data not shown). Interestingly, increased Leu is known to induce LysU expression (125), consistent with the requirement for ProXp-ST1 under these conditions. An analysis of all bacterial genomes reveals that gammaproteobacteria show the largest overlap of lysU and proxp-y genes; 70% of proxp-y-containing genomes also encode lysU, supporting a functional relationship between these two genes in this phylum.

Environmental stress and other factors may result in an imbalance of amino acid levels in the cell, which can pose a difficult challenge for the fidelity of ARSs. In addition to Ser/Thr ambiguity, Hse is a structurally related non-protein amino acid that is
misactivated by LysRS (39) and readily imported into the cell resulting in toxic effects. Although the exact mechanism of Hse toxicity is unknown, Thr can rescue the effect and it is proposed that the toxic effects may be due to misincorporation/misfolding of proteins (126). In addition to LysU, our results showed that Hse is detrimental to the proxp-y null strain in the presence of overexpressed SerRS. However, our in vitro deacylation studies showed that Hse-tRNA\textsubscript{Ser} is not a good substrate for ProXp-ST1 or –ST2, most likely due to the presence of a G discriminator base. We hypothesize that overexpression of SerRS together with Hse may result in mischarging of non-cognate tRNAs in vivo, as previously reported (99). Alternatively, since Hse is an intermediate in the biosynthetic pathway to Thr (126), elevated levels of Hse may result in increased Thr in the cell. Metabolomic studies to directly measure the levels of amino acids inside cells under a variety of conditions will be needed to address these open questions.

In summary, although ProXp-ST1 and –ST2 are homologs of the ProRS editing domain, they do not edit tRNA\textsuperscript{Pro} substrates. These multi-functional domains edit multiple amino acids (Ser, Hse and Thr) and lack significant tRNA specificity, although ProXp-ST2 prefers an A73 discriminator base (data not shown). Thus, this homolog avoids deacylation of aa-tRNA\textsuperscript{Ser}, which encodes a universally conserved G73 in bacteria (127) and both domains fail to deacylate Thr-tRNA\textsuperscript{Thr} in the presence of cognate ThrRS. Promiscuous editing of Hse, a non-protein amino acid, would not be expected to be detrimental to the cell. It has been estimated that nearly 37% of \textit{E. coli} enzymes are “generalists” with multiple substrates, shaped by the metabolic network under certain environmental circumstances (128). Trans-editing proteins such as ProXp-ST1 and
ProXp-ST2 clearly fit into this category, with the capability to act on multiple aa-tRNA substrates depending on the pressures experienced by the cellular translational machinery.

2.5 Acknowledgement

I am grateful to Dr. Oscar Vargas-Rodriguez for the work and discussion on ProXp-ST2 (ProXp-z). We thank Dr. Eva Maria Novoa and Dr. Lluis Ribas de Pouplana for phylogenetic studies and suggestions on the manuscript.
Conserved motifs 1, 2, and 3 (M1-M3, black), the anticodon binding domain (ABD, green), and the editing domain (INS domain, blue) are shown. Single domain INS-like proteins YbaK, ProXp-ala, ProXp-x, ProXp-y, and ProXp-z encoded in the indicated species are shown together with INS and a truncated mini-INS present in the corresponding ProRS. *C. crescentus* is *Caulobacter crescentus* and *R. palustris* is *Rhodopseudomonas palustris*. The known activities of INS, YbaK and ProXp-ala are color coded as follows: blue, Ala deacylation; red, Cys deacylation.
Figure 2.2 Comparison of growth curves of *E. coli* WT (●) and *proXp-y* null strains (○).

*E. coli* WT BW25113 and *proXp-y* null strains (Keio collection) were grown in M9 liquid minimal media containing: (A) 10 mM Ser, AlaRS overexpressed; (B) 10 mM Ser, LysS overexpressed; (C) 10 mM Ser, LysU overexpressed; (D) 1 mM Thr, LysU overexpressed.
Figure 2.3 Effect of Ser, Thr and Hse on the growth of *E. coli* WT (black) and *proXp*-null strains (grey) with LysS, LysU or SerRS overexpression.

*E. coli* WT and proXp-null strains were transformed with overexpression plasmids encoding lysS, lysU or serS genes and cultured at 37 °C overnight. Cells were then washed and inoculated into M9 minimal media contains no additional amino acids, 5 mM of Ser, Thr or Hse. Growth curves were monitored using the Omnilog system. The bars indicate the mean time to ½ maximal O.D. 600 of three independent growth curves. N.G. stands for no growth of cells under the time (48 hours) monitored.
Figure 2.4 Deacylation of 0.75 μM aa-tRNA by 0.5 μM *E. coli* ProXp-y (A, C) or *B. parapertussis* ProXp-z (B, D). All assays were carried out in triplicate with the S.D. indicated.
Figure 2.5 Deacylation of cognate Thr-tRNA$^{Thr}$ by *E. coli* ProXp-y (A) or *B. parapertussis* ProXp-z (B). Reactions were carried out with ~0.75 μM of Thr-tRNA$^{Thr}$ by 0.5 μM ProXp-y or ProXp-z alone or in the presence of 0.3 μM *E. coli* ThrRS or 3 μM *E. coli* ProRS.
Figure 2.6 Deacylation of 0.75 μM Ser- and Hse-tRNA$^{\text{Ser}}$ by 0.5 μM *E. coli* ProXp-y or *B. parapertussis* ProXp-z.
Figure 2.7 Amino acids and tRNA specificity of ProXp-ST1 and ProXp-ST2
<table>
<thead>
<tr>
<th></th>
<th>Concentration</th>
<th>Growth Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Pro</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Cys</td>
<td>0.05 mM</td>
<td>-</td>
</tr>
<tr>
<td>Ser</td>
<td>1 mM</td>
<td>+</td>
</tr>
<tr>
<td>Thr</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Leu</td>
<td>1 mM</td>
<td>+</td>
</tr>
<tr>
<td>Ile</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Gly</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Val</td>
<td>0.005 mM</td>
<td>-</td>
</tr>
<tr>
<td>Lys</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Met</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>His</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Asp</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Asn</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Trp</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Glu</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Gln</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Phe</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Tyr</td>
<td>0.01 mM</td>
<td>-</td>
</tr>
<tr>
<td>Arg</td>
<td>1 mM</td>
<td>-</td>
</tr>
<tr>
<td>Hse</td>
<td>1 mM</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 2.1 Growth curve assays comparing *E. coli* WT and *proxy* null strains
*E. coli* WT and *proxp-y* null strains were first grown in the presence of elevated levels of each of the 20 protein amino acids and homoserine (Hse) in minimal media. For most amino acids, 1 mM was used but due to toxic effects, lower level of Cys, Val and Tyr (0.01-0.05 mM) were used. Comparing the growth curves of the WT strain and the null strain, increased lag phase times were observed in the presence 1 mM Ser, 1 mM Leu and 1mM Hse. “-” stands for the difference in lag phase is less than 10 min, and no growth rate difference is observed. “+” stands for growth defect observed of the null strain.
<table>
<thead>
<tr>
<th></th>
<th>PheRS</th>
<th>HisRS</th>
<th>GlyRS</th>
<th>AlaRS</th>
<th>AsnRS</th>
<th>AspRS</th>
<th>LysS</th>
<th>LysU</th>
<th>SerRS</th>
<th>ThrRS</th>
<th>ProRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser 1mM</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thr 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leu 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ile 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gly 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ala 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Val 0.005mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pro 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lys 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Met 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>His 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asn 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Trp 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glu 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gln 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Phe 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tyr 0.01mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arg 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cys 0.05mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hse 1mM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.2 Growth curve assays comparing *E. coli* WT and *proxp-y* null strains with ARS overexpressed

*E. coli* WT and *proxp-y* null strains were transformed with plasmids encoding each class II *E. coli* ARS. These 22 strains were grown in the presence of elevated levels of each of the same 21 amino acids as in table S1. Differences observed between the growth curves
of the WT and the null strain are summarized as above. “-“ stands for the time difference of WT and the null strain to half maximal O.D. 600 is less than 10 min. “+”, “++” and “+++” mean the difference is below 1 hour, between 1-2 hours and more than 2 hours seperately.
Figure 2.8 Effect of varying concentration of Thr, Ser and Hse on the growth of *E. coli* WT (black) and *proxp-y* null strain (grey) with LysS, LysU or SerRS overexpression is dose-dependent
Chapter 3: YbaK defends against oxidative stress in *E. coli*

3.1 Introduction

Faithful translation of the genetic material is one of the keys to the normal cellular activities. High level of fidelity in translation relies on the correct pairing of codon-anticodon on the ribosome and the attachment of the cognate amino acid onto specific tRNAs by aminoacyl-tRNA synthetases (ARSs). However, the aminoacylation active site of ARS is not always perfect, as it is challenging to differentiate the structure or size of similar amino acids from the cognate one. For example, prolyl-tRNA synthetase (ProRS) mischarges Ala and Cys besides cognate Pro (79). Thus, several proofreading mechanisms have evolved in ARSs to address this problem. In contrast to the pre-transfer editing which happens at the catalytic site, nearly half of the ARSs also use post-transfer editing. This activity is located in an additional editing domain to clear the errors (1). For ProRS, an insertion (INS) domain plays such a role to deacylate the mischarged Ala-tRNA$_{Pro}$ (120). In addition, autonomous editing domain homologs are found throughout genomes. In the ProRS system, YbaK protein, which is a homolog of INS domain, is reported to deacylate Cys-tRNA$_{Pro}$ that escapes from the two active sites of ProRS in vivo (88). A complex of YbaK, ProRS and tRNA$_{Pro}$ is formed to ensure the deacylation specificity (89). YbaK is widely found in bacteria, of 1656 sequenced genomes, over 57 percent encode the *ybak* gene, including pathogenic bacteria such as *Salmonella enterica*.
(Varga-Rodriguez, unpublished data). Here, we investigate the physiological significance of maintaining this discrete protein.

The environments of living cells are not always stable, especially for bacteria. Nutritional imbalance and stress conditions, such as heat shock, cold shock, UV irradiation and oxidative stress, could affect the cell’s metabolism dramatically. Among these factors, oxidative stress triggers complicated cellular reactions and has been studied in a large amount of systems in recent years (129). It has been reported that many stress conditions and pathological states are accompanied by cellular oxidative stress, either as the reason or the result. Oxidative stress is caused by the imbalance between reactive oxygen species (ROS) production and elimination. The main source of endogenous ROS in cells is the electron-transport chains, i.e., the mitochondrial, endoplasmic reticulum, plasma membrane and photosynthetic system (130). Several scavenging systems protect *E. coli* from endogenous ROS, such as superoxide dismutases (SODs), alkyl hydroperoxide reductase (Ahp), catalase G (KatG) and catalase E (KatE) (131). However, when cell faces the condition where large amounts of exogenous ROS exist, the basal defense is not enough. Two systems were discovered as the major regulons for ROS conditions: the OxyR system sensing the level of \( \text{H}_2\text{O}_2 \), and the SoxRS system stimulated by \( \text{O}_2^- \). Besides these two systems, there are additional genes encoding antioxidants up-regulated during oxidative stress (132).

The effect of ROS is profound, including the extensive modification or damage of DNA, RNA, proteins and lipids and induction of mutagenesis. It is commonly believed that this damage would collectively lead to cell dysfunction and death. In fact, oxidative
stress is used by both microbial competitors and eukaryotic hosts to harm bacteria. For example, *E. coli* cells are exposed to H$_2$O$_2$ produced by phagocytes with local concentrations up to 100 µM, or by H$_2$O$_2$-generating lactic acid bacteria with concentration as high as 1 mM (133). Oxidative stress has been shown to affect the accuracy of ARSs in different cell systems. For instance, the editing ability of *E. coli* ThrRS is impaired as the critical Cys residue is oxidized by ROS (98). In mammalian cells, there is evidence indicating that Met is mischarged onto various tRNAs ten times more than normal under condition of oxidative stress (101). It is shown that the misacylation is caused by phosphorylation of MetRS at Ser209 and Ser825 upon ROS stress. The controlled inaccuracy of MetRS defends against the oxidative stress by reducing ROS levels and cell death (134). Considering YbaK is not strictly required for cell growth under normal conditions, it is possible that the freestanding editing protein plays more important role during stress response. In this study, we investigate the behavior of a *ybak* null strain under oxidative stress conditions.

3.2 Materials and Methods

*Materials and strains*—All amino acids and chemicals were purchased from Sigma-Aldrich unless otherwise noted. *Escherichia coli* strain BW25113 (lacI^q^ rrnB_T14 ΔlacZ_WJ16 hsdR514 ΔaraBAD_AH33 ΔrhaBAD_LD78) and the *ybak* null strain (JW0470) from the Keio collection were obtained from the Coli Genetic Stock Center. The gene coding for YbaK was fused to a C-terminal six-histidine extension in the context of plasmid pQE-70 (Amp^r^ marker) (Qiagen, Chatsworth, CA) to generate plasmids pQE-
ybak. Plasmid pQE-ybak was constructed by introducing a SphI site at the 5’ end of the gene and a BamHI site at the 3’ end by site-directed mutagenesis. Mutations of the 5’ TT to GC upstream of the ATG start codon and the sequence at position 2 from a serine codon (AGC) to an arginine codon (GCG) were made to generate a new SphI restriction site (GCATGC). The gene and BamHI restriction site (GGATCC) was obtained from a pet15b vector. The SphI-BamHI restriction fragment containing the ybak gene was purified by standard techniques (87). The fragment was then subcloned into plasmid pQE-70 that was previously digested with the same restriction enzymes. The pQE-ybak plasmid was then transformed into both WT and ybak null strains.

Growth curve assays—M9 minimal media supplemented with 0.4% glucose was used in growth curve assays to culture E. coli WT and ybak null strains. Cells were first grown in LB media overnight, collect and washed twice with minimal media. For strains carrying pQE-ybak plasmid, 100 μg/ml ampicillin was added to rich media. Cells were diluted to an O.D. 600 of 0.01, and grown at 37 °C in M9 minimal media in the presence of Biolog Redox Dye Mix MA, in the absence and presence of 0.1 mM β-D-1-thiogalactopyranoside (IPTG) and the indicate concentration of hydrogen peroxide. Cell growth was monitored at 37 °C for 48 h using the Omnilog® automated system (Biolog Inc., Hayward, CA).

Expression data analysis—Expression data of ybak and related gene under stress condition was obtained from the online databases GenExpDB (http://genexpdb.ou.edu) and PortEco (http://www.porteco.org). All the results are summarized from microarray data.
3.3 Results and Discussion

3.3.1 Expression levels of *proS* and *ybak* gene are regulated during stress.

Transcription levels of genes are subject to changes caused by environmental induced factors. Epistatic regulation of gene expression is critical for maintaining proper cellular activities. The expression of a specific gene could be controlled by more than one transcriptional regulator and suppressed or enhanced depending on the local context. ARSs play a central role in protein biosynthesis and the regulation of ARSs is reported to be related to the growth rate with unknown mechanism in *E. coli*. The expression of most ARSs is induced two- to three-fold for a five-fold increase in growth rate in *E. coli* (135). Adding amino acids to minimal media does not significantly change the concentration of ARSs except threonyl- and phenylalanyl-tRNA synthetases (ThrRS and PheRS). Thus it is quite likely that the expression level of autonomous editing proteins is also regulated under different conditions. In fact, this appears to be the case based on microarray data monitoring YbaK and ProRS expression levels from multiple studies (Fig. 3.1). ProRS levels are mainly down-regulated under cold, heat, oxidative stress and glucose-lactose shift, whereas YbaK levels are consistently up-regulated in at least three stress conditions. The opposite regulation of ProRS and YbaK mRNA level suggests that editing function provided by YbaK could be of special importance under stress conditions.
Figure 3.1 The expression level changes upon extracellular stress. Y-axis: log2 expression ratio of test/control.

3.3.2 H$_2$O$_2$ induces growth defect in $ybak$ null strain.

To investigate the role of YbaK under stress conditions, we tested the growth of an $E. coli$ $ybak$ null strain comparing to the WT strain. Considering inevitable existence of ROS in the cell under aerobic conditions, we focused on the oxidative stress induced by H$_2$O$_2$. H$_2$O$_2$ is permeable to cell membranes (136). In some cases, $E. coli$ could excrete H$_2$O$_2$ to compete with other pathogens by using this limited membrane permeability and expressing high levels of scavenging enzymes in its own cytoplasm. Thus, H$_2$O$_2$ is a naturally occurring source of oxidative stress. In the growth curve assay, 0.1 mM of H$_2$O$_2$ was first used to mimic the host immune environment. As shown in Fig. 3.2A, a slight growth delay of the $ybak$ null strain was observed under these conditions. Upon
increasing the concentration of \( \text{H}_2\text{O}_2 \) to 0.5 mM and 5 mM, more severe growth defects were observed by the null strain (Fig. 3.2B). Specially, when under 5 mM \( \text{H}_2\text{O}_2 \) the ybak null strain could not survive. As the ybak null strain did not shown any growth defect in minimal media under normal growth condition (data not shown), the death of the null strain under severe oxidative stress supports our hypothesis that the physiological role of YbaK is more important in a stress environment.

![Figure 3.2 Growth curve assays for E. coli WT and ybak null strains under different concentrations of \( \text{H}_2\text{O}_2 \).](image)

**3.3.3 Cys amplifies the growth defect under ROS conditions**

Since YbaK’s function is to clear Cys-tRNA\(^\text{Pro} \) both \textit{in vitro} and \textit{in vivo}, we hypothesize that the growth defect of the ybak null strain under oxidative stress conditions is due to the accumulation of mischarged Cys-tRNA species. Addition of Cys to the growth medium is expected to increase mischarging by ProRS, as reported previously (88). As shown in Fig. 3.3A, adding 0.05 mM Cys together with 0.1 mM \( \text{H}_2\text{O}_2 \) indeed induced a larger growth defect compared to 0.1 mM \( \text{H}_2\text{O}_2 \) alone. The lag
time of the ybak null strain under this condition is similar to the delay under 0.5 mM H$_2$O$_2$, which indicates that increasing the Cys level actually amplifies the effect of oxidative stress in E. coli cells. To confirm the defect was due to the missing function of YbaK, we introduced a plasmid encoding the ybak gene and expressed it in both WT and null strains. The overexpression of YbaK could compensate for the defect induced by oxidative stress (Fig. 3.2B). This is consistent with our hypothesis that the function of YbaK becomes significant in a ROS environment.

![Figure 3.3](image.png)

**Figure 3.3** Growth curve assays for E. coli WT and ybak null strain under the indicated concentrations of H$_2$O$_2$ and Cys with (A) or without (B) overexpression of YbaK.

Adding Cys together with H$_2$O$_2$ is expected to result in oxidation of Cys to cystine. How cellular levels of Cys and cystine change under ROS conditions is an open question, as the intracellular concentrations of Cys and cystine are kept extremely low making them difficult to measure directly (137). An interesting finding is that Cys is constantly exported into the periplasm and oxidized to cystine as a reducing agent in E. coli. Under oxidative stress conditions, the genes encoding the exporter of Cys and the importer of cystine are up-regulated. In addition, H$_2$O$_2$ treatment does not promote Cys biosynthesis.
in the cell. Taken together, a Cys/cystine shuttle system strictly controls the redox environment of the periplasm in *E. coli*. Thus, it is unlikely that under ROS conditions, the growth defect of ybak null strains is due to the increasing amount of Cys and consequently elevated Cys-tRNA<sub>Pro</sub>. In addition, although the Cys added together with H<sub>2</sub>O<sub>2</sub> could be oxidized to cystine, the oxidative species will be formed again with the imported of cystine into the cell. In the cytoplasm, the reversible oxidation of Cys occurs. On the other hand, the oxidative environment constantly oxidizes the Cys residue to a Cys-SOH intermediate. Figure 3.4 shows a scheme of Cys biochemistry under oxidative stress (138). Various studies in recent years have confirmed the existence of sulfenic acid modification of protein Cys *in vivo* (132,139,140). It is likely that excess amount of H<sub>2</sub>O<sub>2</sub> alters the composition of the cytoplasmic thiol pool; thus, the accumulation of dysfunctional proteins leads to delayed growth or cell death. The possible role of YbaK in this case is clearing the various forms of mischarged tRNA products, preventing them from entering nascent peptides and protecting the cell from oxidative damage.
Figure 3.4 Intracellular Cys biochemistry. Cys residue could be covalently modified by oxidative stress. The sulfenic form (SOH) is readily reversible, while sulfinic and sulfonic forms are mostly irreversible (138).

3.4 Conclusion

In the case of ThrRS, without knockout of protease, the growth defect of impaired ThrRS is not observed. However, the deletion of the ybak gene alone is sufficient to show a growth defect under oxidative stress conditions. Together with the up-regulation of ybak gene expression, the important role of the freestanding editing domain YbaK in ROS related stress in *E. coli* is revealed for the first time. The opposite regulation of ProRS
and YbaK provides an interesting example showing the advantage of encoding editing
genes separately. When the cells face different stress conditions, especially in the
complex environment of bacteria, controlling the balance between biosynthesis and
quality control processes could benefit the survival of living organisms. In addition, the
study of these protecting mechanisms under stress may provide more therapeutic
strategies for killing pathogenic bacteria.
4.1 Introduction

The correct selection of amino acids by aminoacyl-tRNA synthetases (ARSs) is a key step in translation. However, there are various closely related analogs that could be misactivated, charged onto tRNA and even incorporated into proteins. Early studies had found multiple analogues in polypeptide chains in bacteria, such as ethionine (2-amino-4-(ethylthio) butyric acid), o-, m- and p-fluorophenylalanine, p-aminophenylalanine, 7-azatryptophan, canavanine (2-amino-4-(guanidinoxy)butyric acid) and beta-2-thienylalanine (141-143). Sufficient concentrations of amino acid analogues are required to be incorporated in the cell. The transportation of these analogues is known to use the same transporters as native amino acids, but lower affinities were observed in most cases (144). Beside of structurally related analogues, side-chain modified amino acids, for example, the hydroxyl addition caused by reactive oxygen species, are also candidates for misincorporation (Table 1). Compared to the other analogues, modified natural amino acids are generated intracellularly and released from degraded proteins into the cytosol; thus, they are more likely to access the cellular amino acid pool and be misincorporated into proteins. Essential hydrogen bonds between the modified analogues and ARSs must form for their misincorporation. For example, addition of a hydroxyl usually results in analogues with very similar size and shape, and thus several analogues of this type are
found in proteins. A well-studied case is the abundant 3,4-dihydroxyphenylalanine (DOPA) in tissues of atherosclerotic plaques and the cataract lens (145-147). DOPA is activated by tyrosyl-tRNA synthetase and replaces the Tyr residues in proteins (148-150).

Recent studies have shown that non-protein amino acids are widely found in the human diet, such as seeds, fruits, legumes and nuts. It is proposed that ingestion of these analogues may lead to autoimmune-like symptoms in some cases (151,152). β-N-methylamino-l-alanine (BMAA) is found in cyanobacteria and the seeds of Cycas circinalis (153). A population that consumes this plant in their diet has a 100 fold higher ratio of the occurrence of amyotrophic lateral sclerosis (ALS)/Parkinsonismdementia complex (PDC) (154-156). The relation between BMAA and ALS/PDC patients is also supported by another study in 2009 by Pablo et al (157). In addition to cyrad seeds, BMAA is discovered in shark fins with concentrations from 86 to 265 µg/g (dry weight) (158). In vivo and in vitro studies of misincorporation of BMAA were carried out in mouse and human cells separately (159,160). BMAA can be misincorporated into human proteins in place of Ser, and this process could be blocked by excess Ser. However, the molecular reactions of the mischarging of BMAA by ARSs are unclear. Here, we investigate the mischarging and deacylation of BMAA-tRNA.
Table 4.1 Oxidative modification to amino acid residues found on proteins in vivo (48).

<table>
<thead>
<tr>
<th>Target amino acid residues</th>
<th>Oxidation products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine</td>
<td>Glutamic semialdehyde</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Dithiol, cysteic acid</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Oxalic acid, pyruvic acid</td>
</tr>
<tr>
<td>Histidine</td>
<td>2-oxo-histidine, asparagine, aspartic acid</td>
</tr>
<tr>
<td>Lysine</td>
<td>α-Aminoadipic semialdehyde</td>
</tr>
<tr>
<td>Methionine</td>
<td>Methionine sulfoxide, methionine sulfone</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>α- and m-tyrosine, 2,3-dihydroxyphenylalanine (DOPA)</td>
</tr>
<tr>
<td>Proline</td>
<td>2-Pyrrolidone, 4- and 5-hydroxyproline pyroglutamic acid, glutamic semialdehyde</td>
</tr>
<tr>
<td>Threonine</td>
<td>2-Amino-3-ketoheptonic acid</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>2-, 4-, 5-, 6-, and 7-hydroxytryptophan, 3-hydroxykynurine, nitrotryptophan, formylkynurenine, kynurene</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>DOPA, dityrosine (tyr-tyr cross link), tyrosine-O-tyrosine, crosslinked nitrotyrosine, crosslinked cysteine-DOPA</td>
</tr>
<tr>
<td>Non-aromatic amino acids</td>
<td>Protein carbonyls</td>
</tr>
<tr>
<td>Polypeptide backbone</td>
<td>Protein cross links, alkylperoxide, hydroxyl protein derivatives</td>
</tr>
</tbody>
</table>

4.2 Materials and Methods

Materials—All amino acids and chemicals were purchased from Sigma unless otherwise noted. $[^{32}\text{P}]-\alpha$-ATP was from PerkinElmer Life Sciences.

Enzymes and tRNA preparation—All tRNAs used in this study were prepared by in vitro transcription as described (Beuning and Musier-Forsyth, 2000; Shitivelband and Hou, 2005). His-tagged E. coli ProRS (Wong et al., 2002), WT AlaRS (Beebe et al., 2003), SerRS, LysS, LysU, ThrRS (ASKA collection), E. coli ProXp-ST1, and E. coli tRNA nucleotidyltransferase (NTase) (Nordin and Schimmel, 2002) were purified using the Talon cobalt affinity resin (Clontech). The concentrations of E. coli ARSs other than LysRS, tRNA NTase and all trans-editing proteins were determined by the Bradford assay. The concentrations of E. coli LysS and LysU were determined by active site titration (Fersht et al., 1975).

Aminoacylation assays—3’-[$^{32}\text{P}$]-labeled tRNA for aminoacylation assays was prepared as previously described (Wolfson and Uhlenbeck, 2002). E. coli ARSs 3 μM
were incubated with 3 μM 3'-[^32P]-labeled cognate tRNA and 1 mM BMAA in 50 mM HEPES (pH 7.5), 4 mM ATP, 20 mM KCl, 50 mM dithiothreitol, 25 mM MgCl\(_2\), and 0.1 mg/ml bovine serum albumin (BSA) at 37 °C. For each time point, 1.5 μL of reaction mixture was quenched into 4 μL of 200 mM NaOAc containing 0.4 unit/μL P1 nuclease at 4 °C. After digestion for 20 min at room temperature, 1 μL of quenched mixture was spotted onto a polyethyleneimine-cellulose TLC plate pre-run with water. Separation of BMAA-AMP and AMP was accomplished by developing the TLC in 0.1 M ammonium acetate and 5% acetic acid. The radioactivity was analyzed by using a Typhoon phosphorimager and data were analyzed using Biorad Quantity One software. The fraction of free tRNA was determined from the ratio of AMP formed over the total AMP plus BMAA-AMP. Preparation of BMAA-tRNAs by ARSs and aa-tRNA-Lys by flexizyme for use in deacylation assays was carried out as previously described.

**Deacylation assays**—In vitro deacylation assays were performed at 37 °C in reactions containing ~5 μM[^32P]-labeled aa-tRNA, 50 mM HEPES (pH 7.5), 2 mM dithiothreitol, 20 mM KCl, 5 mM MgCl\(_2\), 0.1 mg/ml BSA and 15 μg/ml inorganic pyrophosphatase. Reactions were initiated by addition of 1 μM *E. coli* ProXp-ST1, and quenched at indicated time points by putting 1.5 μl of the reaction into 4 μl of a solution containing 200 mM NaOAc (pH 5.5) and 0.4 U/μl of P1 nuclease.

4.3 Results and Discussion

4.3.1 BMAA is mischarged by SerRS, LysRS, AlaRS and ThrRS
As shown in Fig. 4.1A, the structure of BMAA is similar to several amino acids such as Ser and Thr, and thus is potentially capable to be mischarged by various ARSs. *In vitro* aminoacylation assays were carried out at 37 °C with the same concentration of substrates and reagents. As shown in Fig. 4.2, except for ProRS, all ARSs tested are able to mischarge BMAA onto their cognate tRNAs. The highest level of mischarging was observed for SerRS consistent with the *in vivo* replacement of Ser by BMAA. Though mischarging of BMAA by other ARSs has not been explored *in vivo* to date, it is possible that BMAA is misincorporated at multiple sites in proteins.

Figure 4.1 Structure of BMAA (left), Ser (middle) and Thr (right).
4.3.2 BMAA-tRNA could be deacylated by ProXp-ST1

As reported in Chapter 2, ProXp-ST1 is proposed to be a freestanding editing protein that deacylate Ser/Thr-tRNA formed by mischarging error caused by multiple ARSs such as AlaRS, ThrRS, LysRS and SerRS. We hypothesize that ProXp-ST1 is able to deacylate the mischarged BMAA-tRNAs. As shown in Fig. 4.3, various BMAA-tRNAs are deacylated by ProXp-ST1 to different extents. As the deacylation reactions were carried out at 37 °C, the buffer hydrolysis of BMAA-tRNA itself is dramatic. In addition, the deacylation rates of BMAA-tRNA by ProXp-ST1 are not the same and appear to depend on the tRNA species. Interestingly, BMAA-tRNA\textsuperscript{Ser} and BMAA-tRNA\textsuperscript{Thr} are the most robust substrates for ProXp-ST1.
4.3.3 P5C is mischarged by ProRS

Pro metabolic pathway is reported to be involved in stress defense responses including scavenging of ROS (161). Pro is oxidized by two dehydrogenases to form glutamate, wherein pyrroline-5-carboxylate (P5C) is formed during this pathway (Fig. 4.4). Mischarging of P5C by *E. coli* ProRS was performed in vitro (Fig. 4.5). The mischarging of this intermediate suggests that the aminoacylation site of ProRS is capable of accepting P5C as substrate.
4.3.4 EtSer- and iPrSer-tRNA could be deacylated by ProXp-ST1

Several synthesized analogues of Ser and Thr were obtained and mischarged onto *E. coli* tRNA$_{Lys}$ by flexizyme (Fig. 4.6). As shown in Fig. 4.7, EtSer- and iPrSer-tRNA$_{Lys}$
are robustly deacylated by ProXp-ST1, while Dap- and D-Thr-tRNA^{Lys} are not. These results suggest the active site of ProXp-ST1 could accommodate larger substrates than Thr.

Figure 4.6 Structures of Ser/Thr analogues linked with leaving groups in flexizyme reactions.
4.4 Conclusions

Increasing number of non-protein amino acids that are misincorporated by ARSs have been reported in recent studies. In addition to possible disease-related consequences, these findings on structurally similar substrates expand the genetic code and may have positive applications in protein engineering. In the case of BMAA, the detailed molecular mechanism of mischarging and the cellular effects are still unknown. Future work could include kinetic studies of aminoacylation by various ARSs and the in vivo effects of BMAA in different cell systems.

Figure 4.7 Deacylation of aa-tRNA^{lys} by ProXp-ST1.
5.1 Probing the \textit{in vivo} interacting partners of ProXp-ST1 and YbaK

ProXp-ST1 is proposed to interact with various ARSs to clear the misacylated aa-tRNAs. For YbaK, as Cys is misactivated by several ARSs such as IleRS, ValRS, AlaRS, LysU and SerRS in addition to ProRS, interaction between YbaK and these ARS is also possible. Exploring the interacting partner of YbaK and ProXp-ST1 would help address the real substrate of these two proteins \textit{in vivo}. Crosslinking by UV (CLIP) and formaldehyde methods are used to trap the temporary interactions between freestanding proteins and potential interacting tRNA and ARS by linking them covalently (Fig. 5.1). Polyclonal antibody of ProXp-ST1 and YbaK are utilized to immunoprecipitate the protein-tRNA complex. Before the incubation with cell lysis, antibodies were further purified from serum by coupling of the antigen to cyanogen bromide activated sepharose following a standard protocol. Following gel electrophoresis and membrane transfer ensures the separation of protein-tRNA complex of correct size. After digestion by proteinase K, the remaining RNA sample is ready to be used for sequencing.
Figure 5.1 A schematic representation of CLIP procedure for indentification of protein binding sites on RNAs in the context of intact cells (162).

The next step involves a group II intron reverse transcriptase (RTs) used to introduce an adaptor to the end of RNA samples (163). The group II intron RTs with template switching capability could efficiently and seamlessly link PCR primer binding sites to cDNA ends without an RNA ligase step (Fig. 5.2). Subsequent sequencing of RNAs will be performed by cloning the fragments into TOPO vectors. Once reliable results are obtained, more conditions, such as oxidative stress and different nutrients, could be applied to the growth of *E. coli* cells to investigate the substrate dependence on the environment. In addition, the protein interaction partner could be recognized from the tRNA-protein complex by mass spectrometry.
5.2 Kinetic study of ARSs with non-cognate amino acids under stress condition

The enzymatic activities of ARSs could be subject to change under different conditions. The reason that *E. coli* maintains two isoform of LysRS in its genome is unclear. Measuring the parameters for activation and aminoacylation of Ser and Thr by LysS and LysU under different temperature is helpful in understanding the role of ProXp-ST1 in LysRS system. Except for heat shock, how oxidative stress affects the aminoacylation and editing activities of ARSs is largely unknown. Especially, since ProRS mischarges Cys, which is a main target of ROS in protein, examination of the editing activity of the modified amino acids would assist drawing the whole picture of the triple sieve mechanism of ProRS system. Besides, like ThrRS, AlaRS also employ a functional Cys residue in its editing site, and could be subjected to modification under stress condition similarly.
5.3 X-ray crystallography studies of ProXp-ST1

Though structures of various INS domain homologs are solved (INS, PDB 2J3L; YbaK, PDB 2DXA, 1DBU; ProXp-ala, PDB 1VK1, 1VJF; ProXp-x, PDB 2CX5 and 1WDV), the tertiary structure of ProXp-ST1 is not available yet. Considering the unique substrate specificity, obtaining crystals of ProXp-ST1 together with its substrate could provide detailed active site structure and information about critical residues and their interactions with the substrate. During current crystal screen set-ups of ProXp-ST1 alone, two conditions were found (Table 5.1). However, further optimization by hanging drop method failed to crystalize. For future work, in spite of following optimization of ProXp-ST1 alone, a non-hydrolyzable Ser- or Thr-tRNA could be synthesized and co-crytalized with ProXp-ST1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Precipitation Reagent</th>
<th>Buffer</th>
<th>Salt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10% (v/v) 2-propanol</td>
<td>100 mM Sodium phosphate dibasic/citric acid pH 4.2</td>
<td>200 mM lithium sulfate</td>
</tr>
<tr>
<td>2</td>
<td>30% (v/v) PEG 400</td>
<td>100 mM Sodium cacodylate/hydrochloric acid pH 6.5</td>
<td>200 mM lithium sulfate</td>
</tr>
</tbody>
</table>

Table 5.1 Buffer conditions from Wizard screen kit for ProXp-ST1 crystals
5.4 Metabolomic stress response

The levels of amino acids and metabolites in *E. coli* are subject to change under environmental fluctuations. While microarray studies of expression level of ARSs and homologs only provide the regulation on the level of transcription, the perturbation of the level of substrates of ARSs, i.e., amino acids and metabolites, under different stress conditions is not fully understood. For example, at 42 degree, comparing to WT strain, *proXp-st1* null strain showed an enlarged growth defect under 10 mM Thr condition when LysU is overexpressed (Fig. 5.3). This result suggests the role of ProXp-ST1 is important when LysU is induced. However, to confirm the growth defect is caused by the elevated mischarging ratio of Thr-tRNA\textsubscript{Lys}, we need to first check the Thr is transported and elevated under these conditions. Metabolomic studies will be carried out for *E. coli* WT and *proXp-st1* null strains under conditions of heat, cold, oxidative stress and under elevated level of different amino acids. The level of other metabolites, such as homoserine, will also be examined to understand the changes of levels of non-proteinogenic substrates of ARSs and their potential misincorporation into proteins.
Figure 5.3 Growth curves of *E. coli* WT and proXp-st1 null strain at 42 degree under Thr condition when LysU is overexpressed.
References


studies of its interaction with tRNA(Lys). *J Mol Biol* **253**, 100-113


60. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S.,


84. Guo, M., Chong, Y. E., Beebe, K., Shapiro, R., Yang, X. L., and Schimmel, P. (2009) The C-Ala domain brings together editing and aminoacylation functions on one tRNA. Science 325, 744-747


97. Lee, J. W., Beebe, K., Nangle, L. A., Jang, J., Longo-Guess, C. M., Cook, S. A.,


10494-10506


for amino acid editing activity. *Biochemistry* **46**, 6258-6267


Escherichia coli. Proc Natl Acad Sci U S A 102, 9317-9322


acid synthetases from *Escherichia coli* and Bacillus subtilis. *Biochemistry* **5**, 1690-1695


