CHARACTERIZATION OF A TRYPTOPHANYL-TRNA SYNTHETASE URZYME AND ITS INTRAMOLECULAR COMPLEMENTATION ESTABLISHES A MODEL TO STUDY THE CATALYTIC MECHANISM AND EVOLUTION OF THE CONTEMPORARY CLASS I AMINOACYL-TRNA SYNTHETASES

Yen Bao Pham

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics, School of Medicine.

Chapel Hill
2009

Approved by:

Advisor: Charles W. Carter, Jr.

Committee member: Howard M. Fried

Committee member: Brian Kuhlman

Committee member: Aziz Sancar

Committee member: Richard Wolfenden

ABSTRACT

Yen Bao Pham: Characterization of a tryptophanyl-tRNA synthetase Urzyme and its intramolecular complementation establishes a model to study the catalytic mechanism and evolution of the contemporary Class I aminoacyl-tRNA synthetases
(Under the direction of Charles W. Carter, Jr.)

Aminoacyl-tRNA synthetases (AaRSs) comprise a crucial group of enzymes catalyzing a two-step reaction to activate amino acids and acylate their cognate tRNAs. It is widely known that highly conserved sequences and motifs divide the synthetases evenly into two distinct classes of ten members each. However, the questions of how they might have emerged and evolved and what a reasonable candidate for the ancestral enzyme would be like still remain unsolved. Answering these would not only shed light on the aaRSs evolution but could also help unravel the critical point when the extremely high kinetic barrier of amino acid activation was bypassed, allowing the synthesis of primitive peptide chains.

The research object of the studies is a tryptophanyl-tRNA synthetase (TrpRS) Urzyme (Ur-: prefix means primitive, original) which would fit the description of a potential candidate for the ancestral class I aaRS. The TrpRS Urzyme is designed to retain class I signature sequences that comprise the ATP-binding Rossmann fold. Interestingly, it has similar affinity for ATP compared to the full length wild type enzyme but surprisingly low affinity for the cognate amino acid substrate, tryptophan. This might be explained by absence of the connecting peptide 1 (CP1), an insertion that wraps around the tryptophan binding site in the contemporary enzyme and occurs at the same location as the editing domain in four other class I aaRSs (MetRS, LeuRS, IleRS, ValRS). An *in cis* modular complementation experiment might point out a

potential role for CP1 in configuring a confirmation that has high affinity for tryptophan, functionally similar to a primitive proofreading function. Furthermore, an Urzyme containing the C-terminal domain (CTD) *in cis* has been constructed to further explore CTD's role in tryptophan activation and $tRNA^{Trp}$ acylation.

The active TrpRS Urzyme establishment confirms an essential prediction of, and is therefore, the first experiment to provide evidential support for the unique Rodin-Ohno hypothesis about the earlier existence of a sense-antisense coding system for both aaRS classes. The successful design of the TrpRS Urzyme has also provided a precedent that has led to the construction of class I and class II aaRS Urzymes of comparable lengths.

ACKNOWLEDGEMENTS

To accomplish the required work for the doctoral degree, I would like to thank my advisor, Dr. Charles W. Carter, Jr. for his tremendous support, encouragement, inspiration and advice. I am also grateful to have four committee members who have been very helpful keeping me on the right track as well as contributing to various projects crucial to the completion of my dissertation.

Dr. Fried has been a supportive and understanding committee chair constantly giving new ideas and suggestion to solve occurring problems. In the same way, the computational phase of my design was accomplished with the guidance from Dr. Kuhlman. Equally important, necessary reagents for my experiments were kindly and generously provided by Dr. Sancar. Last but not least, Dr. Wolfenden has suggested the use of pyridine in the pyrophosphate exchange assay which is critical to measuring the weak activity of the TrpRS Urzyme. Members of all labs involved were of great help when the details for experiments were sought.

Next, I would like to express my regards to the Vietnam Education Foundation (VEF) for awarding me a scholarship to pursue the doctoral degree at the University of North Carolina at Chapel Hill. The foundation had fully funded my research for the first four years and partially supported the fifth year.

Finally, I would like to thank my parents and my brother, who are in Vietnam, our home country. They cannot physically stay beside me, however, have continuously shared important moments, happy or sad. Our regular conversations are always motivating and encouraging so that I would be able to complete my goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aa	Amino acid
AaRS	Aminoacyl-tRNA synthetase
ABC	TrpRS monomer with truncated CTD
AC	Fragment constructed by fusing two halves of the Rossmann fold
ACD	TrpRS monomer with deleted CP1
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
B	Connecting peptide 1 (CP1)
CP1	Connecting peptide 1
CTD	Carboxyl-terminal domain
D	Carboxyl-terminal domain (CTD)
DES	Redesigned TrpRS Urzyme
HIGH	Sequence composed of histidine (H), isoleucine (I), and glycine (G)
KMSKS	Sequence composed of lysine (K), methionine (M), and serine (S)
MBP	Maltose binding protein
MSA	Multiple sequence alignments
ORF	Open reading frame
PP_i	Pyrophosphate
SASA	Solvent accessible surface area
SUMO	Small ubiquitin-related modifiers
TEV	Tobacco Etch Virus
TrpRS	Tryptophanyl-tRNA synthetase
WHEP	A conserved domain found in TrpRS (W), HisRS (H) and Glu-ProRS (EP)
WT	Wild type TrpRS Urzyme

CHAPTER I

INTRODUCTION

Aminoacyl-tRNA Synthetases Play a Crucial Role in Various Biological Processes

The major function of these enzymes, as suggested by their names, is to synthesize specific aminoacyl-tRNAs (II) (Carter, 1993). The first half of this reaction, amino acid activation by ATP (I), is the rate limiting step among those required for protein synthesis on the ribosome since this is the only means whereby the free amino acids can be activated and linked with their cognate carriers. It is important to note that (I) occurs at a slow rate of 8.3 x 10^{-9}/mol/s (as corrected from the published value of 7 x 10^{-6}/mol/min for reaction at pH 9.7, 39°C for a similar reaction (Kirby and Younas 1970) in the absence of this biological catalyst due to its high activation energy barrier. Hence, the emergence of aaRS enzymes is a very significant step in launching the evolution of protein synthesis.

$$\text{aaRS} + \text{ATP} + \text{aa} \Rightarrow \text{aa-AMP-aaRS} + \text{PP}_i \quad \text{(I)}$$

$$\text{aa-AMP-aaRS} + \text{tRNA} \Rightarrow \text{aa-tRNA} + \text{AMP} + \text{aaRS} \quad \text{(II)}$$

An intriguing characteristic of aaRS is that they can be either accurate or ambiguous depending on the enzyme's ability to recognize its cognate substrates, or in other words, its specificity. There exists an internal domain called the editing domain in some aaRSs that corrects the formation of misacylated amino acids through hydrolysis, thus further ensuring the fidelity of genetic code in translation. Editing-defective aaRSs have been shown to cause unfaithful translation, misfolding and neurodegeneration (Lee *et al*., 2006). On the other hand, the flexibility of the same enzyme provides materials for evolutionary changes in protein engineering (Link *et al*., 2003). The introduction of different amino acids other than the twenty limited to natural proteins by a promiscuous aaRS is the major step in expanding the current genetic code. Two non-canonical amino acids, selenocysteine and pyrrolysine, coined the 21st and 22nd amino acid, respectively, have been found encoded by codons once

thought to be non-sense, stop triplets. Associated with these new amino acids are aaRSs that have evolved respectively, from seryl and lysyl tRNA synthetases (Zhang *et al*., 2005).

As summarized in Figure 1, beside their aminoacylation activity, aaRSs have been recently linked to a number of biological processes (Park *et al*., 2006). These include signaling pathways in human, in which the Class Ic tryptophanyl and tyrosyl synthetases play angiogenic and antiangiogenic roles, respectively (Wakasugi *et al*., 2002; Yang, 2004; Tzima *et al*., 2005). More interestingly, the assembly of different aaRSs in a multi-aaRS complex (MSC) has been shown to be involved in cell death (apoptosis) (Hausmann and Ibba, 2008). In a different biological process, studies by Fox *et al.* (2008) pointed out that the WHEP domain of glutamyl-prolyl-tRNA synthetase (EPRS) plays a dual role in inflammation-responsive, transcript-selective translational silencing. Importantly, according to Musier-Forsyth *et al.* (2006), HIV virion assembly requires formation of a packaging complex which contains $tRNA^{Lys}$, substrate of lysyl-tRNA synthetase (LysRS), and the synthetase itself. Implications of aaRSs in a variety of diseases, including cancer, Alzheimer's (Paley *et al*., 2007), Charcot Marie-Tooth neuropathy (Seburn *et al*., 2006) make understanding both the aaRSs structure and the mechanism a significant object for the pharmaceutical industry.

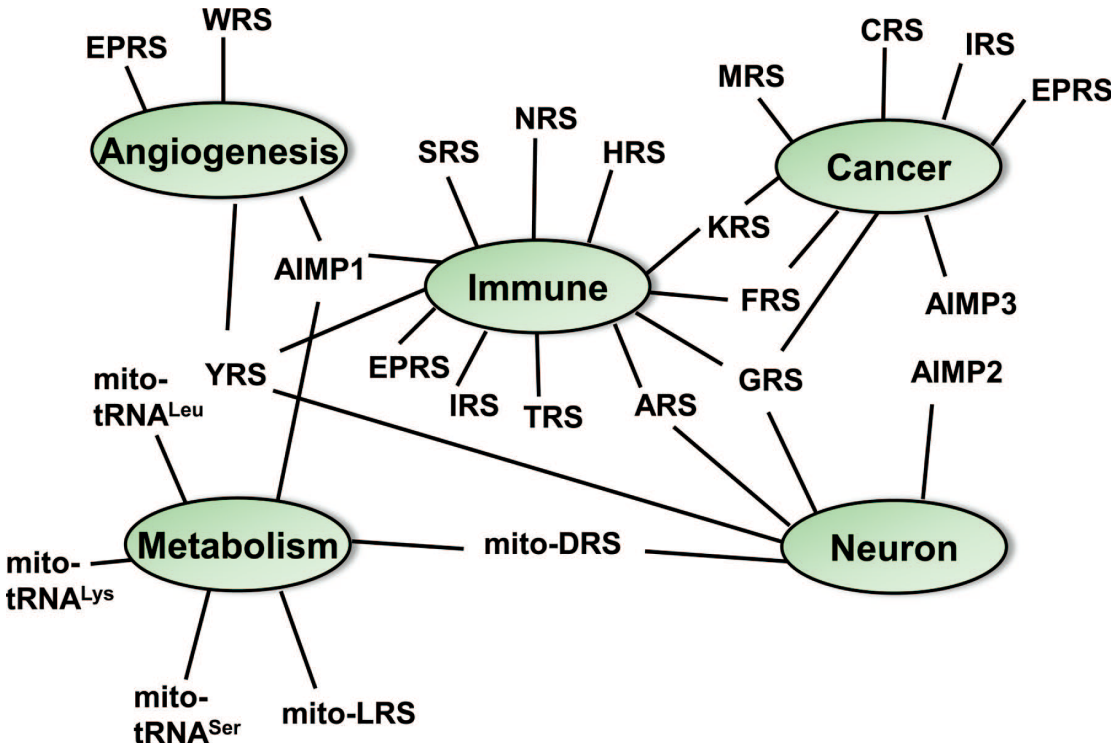


Figure 1: **Connections of aaRSs in various biological processes showing their potential in medical treatments** (Park *et al*., 2006).

My dissertation aims to answer two major questions, the first regarding aaRS modular complementation (intramolecular interactions), and the second about its evolution by constructing a novel tryptophanyl-tRNA synthetase *Urzyme* (Ur-: suffix for primitive, original). As the name suggests, an Urzyme is a construct designed to represent an early stage in the evolution of a contemporary enzyme. It consists of a minimal catalytic domain with sufficient stability to fold and which retains enough of the active site to show a considerable fraction of the catalytic activity of the native protein.

Structural Composition of Contemporary AaRS and Its Classification

Based on structure and catalytic mechanism, twenty contemporary aaRSs are classified into either class I or class II, each consisting of ten members. Class I aaRSs are defined by their two signature sequences of Lys-Met-Ser-Lys-Ser (KMSKS) and His-Iso-Gly-His (HIGH) while three conserved regions, motifs 1, 2 and 3 make class II's identity. Further studies divide each class into three subclasses a, b and c (Eriani *et al*., 1990; Carter, 1993; Cavarelli *et al*., 1994). The structures of the members in each subclass are similar, judging from superimposition. The non-random homology in structures and mechanism raises an unsolved question of how the enzymes might have evolved. The question is even more complicated upon factoring in the existence of the internal and external insertions in between the highly conserved sequences and motifs.

The Emergence of Aminoacyl-tRNA Synthetases

Molecular evolution is a field that draws a lot of attention. The evolutionary standpoint allows researchers to investigate the past, while also opening doors to the future through modeling and prediction (Dean and Thornton, 2007). It is widely accepted that primitive proteins were small and simply structured, and that they

gradually acquired multi-domain, multi-function form through evolution. Catalytically active proteins, enzymes, have probably evolved most profoundly under the pressure to be a more efficient catalyst. This evolution could be achieved by various means, e.g. gene duplication, domain fusion, substrate binding pocket reconfiguration.

The aaRSs are a dynamic class of enzyme. One can see the emergence of new ways to increase aminoacylation activity in living organisms. The activity of one aaRS might increase several times in the presence of another (Prætorius-Ibba, Rogers *et al*., 2005). In a similar way, different fusion forms of aaRSs have been found to have dual activity (Sampath *et al*., 2004). Furthermore, some aaRSs have been shown to be associated with an elongation factor at the site of ribosomal protein synthesis which would enhance the efficiency of the process (Hausmann and Ibba, 2007). These are the results of recent studies on different aaRSs focusing on intermolecular interactions. However, the intramolecular interactions among aaRS structural domains still raise many unresolved questions that need to be answered to unravel the concerted catalytic mechanism.

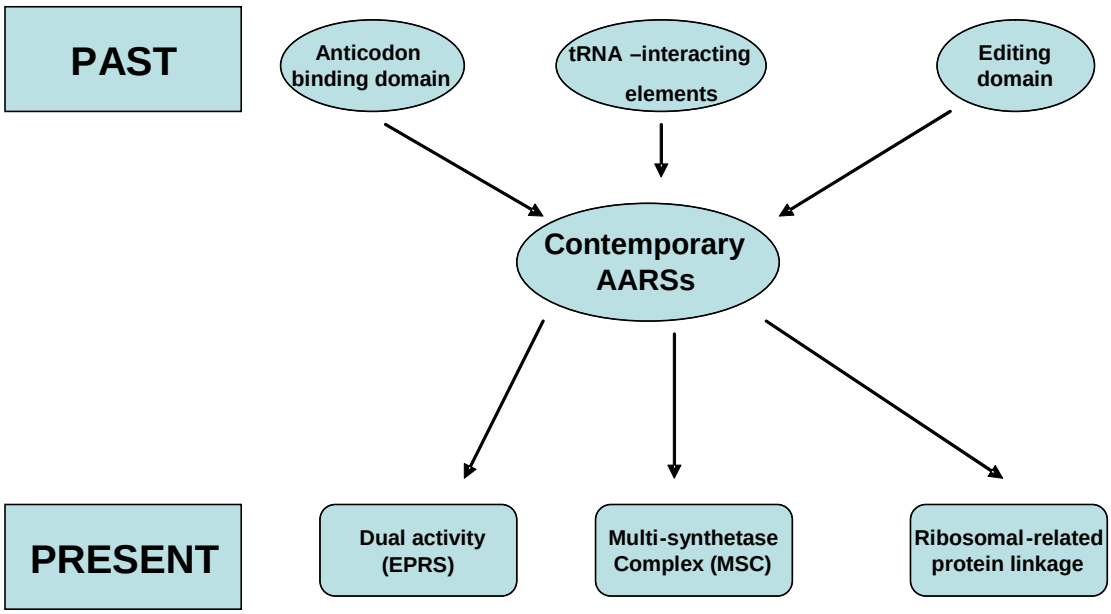


Figure 2: **Postulated evolution scheme of aminoacyl-tRNA synthetase from a primitive form to the contemporary complex fusions** (Pham, unpublished).

Besides the continuing dynamic changes, there is evidence about the insertions of functional domains from a distant past. Larger aaRSs possess different internal and external insertions. These insertions include the connecting peptides 1 and 2, the anticodon binding domain and other structures of unknown functions (Burbaum, Starzyk, and Schimmel, 1990). Connecting peptide 1 probably is one of the best characterized insertions. It has been shown that the editing domain in four contemporary class I aaRSs (MetRS, LeuRS, IleRS,ValRS) probably originated by elaboration from the connecting peptide 1 insertion. The carboxyl-terminal anticodon binding domain and other RNA-interacting elements were also likely introduced later in evolution. Some groups have dissected parts of these domains to construct minimal aaRSs (Schwob and Soll, 1993; Augustine and Francklyn, 1997). However, no study has been done to find out how ancestral catalysts might have emerged from the presumably small and simpler peptides. A postulated scheme for aaRS evolution is depicted in Figure 2.

This question becomes more interesting and more complicated since there are two different classes of contemporary aaRSs. Would they be the products of divergent or convergent evolution? Rodin and Ohno have postulated a unique hypothesis regarding the emergence of aaRSs from two strands of one ancestral gene (Figure 3).

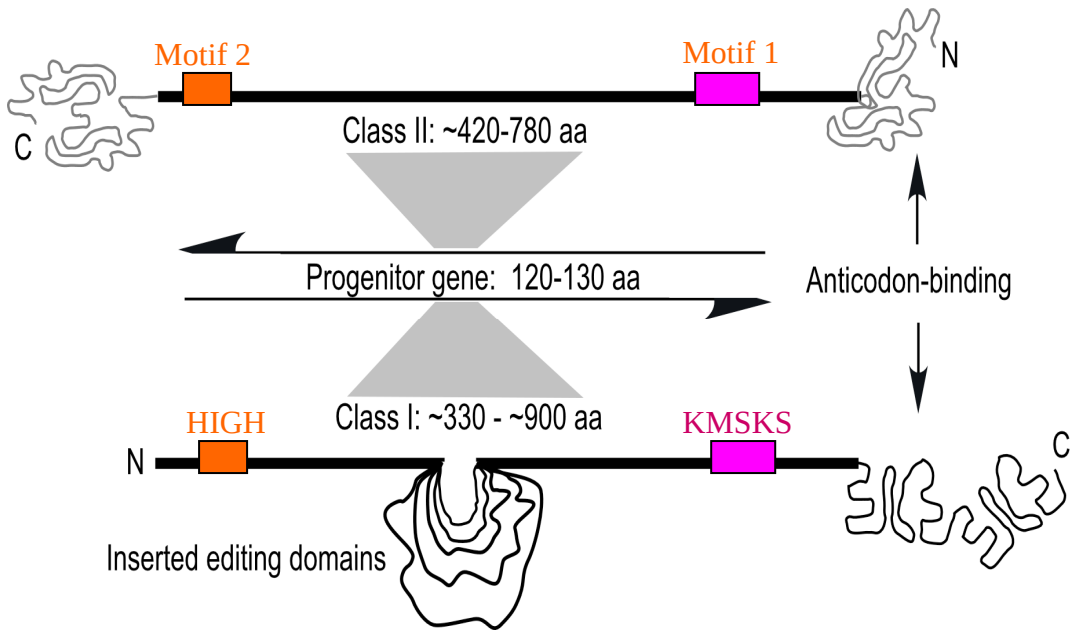


Figure 3: **Alignment of class I and class II aaRSs most highly conserved motifs and sequences showing a sense-antisense homology relationship as an illustration for Rodin-Ohno hypothesis** (adapted from Pham et al., 2007).

This work on the TrpRS Urzyme confirms a key prediction of the Rodin-Ohno hypothesis (Rodin and Ohno, 1995) that ancestral Class I enzymes must have been functional without the CP1 insertion. As such, it provides the first experimental evidence to support that hypothesis and introduces procedures that may be useful to characterize and study the Urzymes of other multi-domain proteins. As shown in Figure 3, the CP1 insertion (corresponding to the editing domain in four aaRSs) and the C-terminal anticodon-binding domain were removed to reveal a minimal catalytic core that comprises the TrpRS Urzyme. The computational construction phase and the significant activity from a refolded TrpRS Urzyme will be discussed in the second chapter. The third chapter is focused on the investigation of a suitable expression system to solubilize the Urzyme. Using an optimized system, Michaelis-Menten kinetics for the Urzyme are characterized and will be presented in the fourth chapter.

Modular Complementation Starting with a Core Urzyme to Understand the Intramolecular Interactions

I have made two additional constructs, in which the connecting peptide 1 and the C-terminal anticodon-binding domain are located *in cis* relative to the core TrpRS Urzyme. An immediate future direction of the dissertation is to investigate the role of those two domains in amino acid activation, specificity and acylation of $tRNA^{Trp}$. These studies are intended to support the idea that the contemporary enzyme may once have existed as a simple peptide chain that acquired additional functional domains through later insertions in evolution.

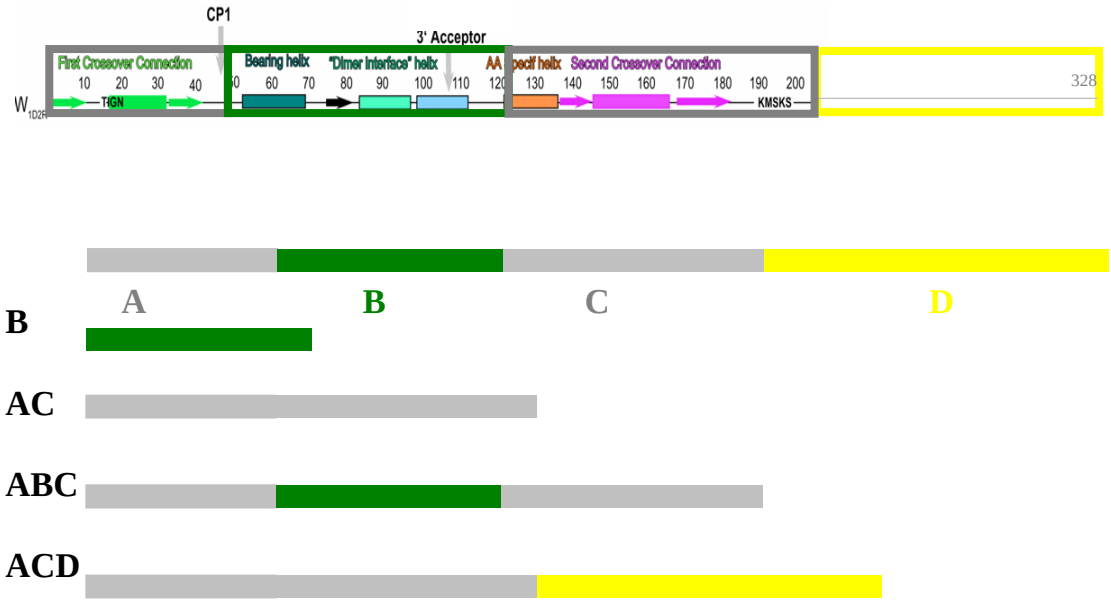


Figure 4: Modular fragments designed to examine the complementation of TrpRS Urzyme. A, C: two discrete halves of the Rossmann ATP binding domain (colored gray). B: connecting peptide 1 (green), a 74-residue long peptide insertion between A and C fragments. D: C-terminal anticodon binding domain (yellow). ABC, ACD: constructs containing TrpRS Urzyme and B and D *in cis*, respectively. (Pham, unpublished).

CHAPTER 2

A RE-DESIGNED, CATALYTICALLY ACTIVE TRYPTOPHANYL-TRNA SYNTHETASE URZYME LACKING THE ANTICODON-BINDING DOMAIN AND AN INTERNAL 76-RESIDUE PEPTIDE PROVIDES THE FIRST EXPERIMENTAL EVIDENCE FOR THE SENSE/ANTISENSE CODING ANCESTRY OF CLASS I AND CLASS II AMINOACYL-TRNA SYNTHETASES

Abstract

Every protein starts from the translation of codons in messenger RNA into a chain of amino acids. There are twenty contemporary aminoacyl-tRNA synthetases (AaRS) initiating the assembly of polypeptide chains by catalyzing the formation of covalent cognate amino acid – transfer RNA adducts so that the protein building blocks can be carried to the ribosomal machinery for protein synthesis. It is well established that these enzymes belong to two different superfamilies called class I and class II aaRS, each consisting of 10 members. Tryptophanyl-tRNA synthetase is the smallest member of class I aaRS, possessing two signature sequences TIGN and KMSKS that form its active site. Interestingly, this active site is separated by a connecting peptide (CP1) that, when removed *in silico* using computational design, leaves a gap of only 5 Ångstrom. Using computational algorithms, a stable minimal catalytic domain, designated the TrpRS Urzyme (Ur- : prefix means primitive, original), has been designed by fusing the two active site segments after removing CP1 and the C-terminal anticodon binding domain (CTD). Pyrophosphate exchange assays have showed significant amino acid activation activity for the molecule whose size is only one sixth that of the active dimer. TrpRS Urzyme may have evolved by acquiring other domains later in evolution. Further experiments are to be done to validate this minimal catalytic domain as a candidate for an ancestral TrpRS gene (this chapter is adapted from Pham et al., 2007).

Introduction

Proteins, especially enzymes, have been intensively studied since the beginning of life sciences due to their universal participation in every biological process. A widely accepted common model for protein synthesis by ribosomes has been well established. However, the evolution of the polypeptide synthesizing machinery is still an unresolved area. One prevailing hypothesis postulates the existence of an "RNA world". Supporters of the RNA world hypothesis are struggling to find RNA capable of amino acid activation. This reaction is kinetically very slow, or in other words, needs a great amount of energy to reach transition state (Kumar and Yarus, 2001). The activated amino acid is by definition a reactive species and hence reacts rapidly, about 10 million times more than amino acid activation. Thus, peptide bond formation from activated amino acids proceeds rather fast compared to amino acid activation by ATP, which is very slow, and the actual rate limiting step in protein synthesis is the activation of the amino acids. Acylation of tRNA is also much more rapid than amino acid activation. Hence, aminoacyl-tRNA synthetases, the contemporary enzymes catalyzing this difficult task, are the most important missing catalysts in the big picture linking the postulated primitive RNA world and the present proteome.

Aminoacyl-tRNA synthetases are the solution for rate enhancement in living organisms. The crucial role of AaRSs in protein synthesis is comparable with DNA polymerase in replication (Francklyn, 2008). Polymerases have a large rate enhancement (i.e., they are very good catalysts) and they also ensure faithful copying of the genetic messages, both in replication and transcription. Similarly, in addition to accelerating the rate of amino acid activation, aaRSs also recognize both their cognate amino acids and tRNAs. It has been shown that five aaRSs possess an editing capability that enhances the fidelity of translation. Since protein synthesis is common

to every organism, aaRSs can be used to construct a phylogenetic tree. New codon translations are introduced when this proofreading activity is compromised.

AaRSs use energy from adenosine triphosphate in a two-step reaction to activate free amino acids (I) then acylate the cognate tRNAs (II) to form aminoacyl-tRNAs for protein synthesis on the ribosome.

$$\text{aaRS} + \text{ATP} + \text{aa} \Rightarrow \text{aa-AMP-aaRS} + \text{PP}_i \qquad \text{(I)}$$

$$\text{aa-AMP-aaRS} + \text{tRNA} \Rightarrow \text{aa-tRNA} + \text{AMP} + \text{aaRS} \qquad \text{(II)}$$

The twenty canonical aaRSs are divided into two classes, depending on structural and mechanistic characteristics. Ten members in class I aaRS have two signature loops, KMSKS and HIGH, which are parts of the Rossmann fold (whose core is a parallel β-sheet flanked by α-helices) responsible for the ATP and tRNA acceptor stem binding. On the other hand, ATP and tRNA interaction sites of class II aaRSs that form the anti-parallel β-sheet core consist of different combinations of three degenerate sequence motifs. Among those, motif 2, ΦΦΦ±ΦxxxFRxE/D (Φ: hydrophobic residue), is shared by all members (Carter, 1993). Based on structure and reaction mechanism, ten aaRSs in each class are also divided into subclasses a, b and c, with subclass a being the largest, containing aaRSs for six amino acids in each case, and subclass c being the smallest, containing aaRSs for the aromatic amino acids. Interestingly, the catalytic cores of class I and class II aaRSs shown in Figure 5 consist of only 120-130 amino acids.

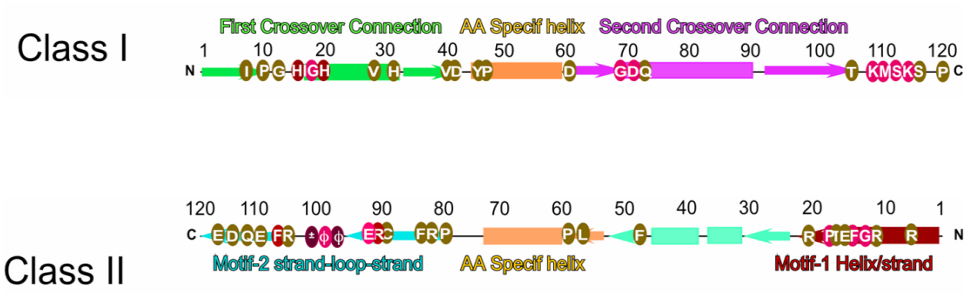


Figure 5: The cores of Class I and Class II aminoacyl-tRNA synthetases. Approximately 120-130-residue long peptides carved out from conserved class I signatures HIGH and KMSKS and class I motifs 2 and 1, then aligned in opposite directions (Carter lab, unpublished).

Most class I aaRSs bind the tRNA acceptor stem from the minor groove side and acylate the 2'-OH group of the terminal base whereas most class II aaRSs bind to the major groove side and acylate the 3'-OH group (Eriani *et al*., 1990; Carter, 1993; Ribas de Pouplana and Schimmel, 2001). One common feature of class I enzymes is that their active form is usually monomeric. In contrast, most class II enzymes are active only in oligomeric form and in which motif 1 contributes essentially to a dimer interface. The consistency within each class and the distinction between the two classes make the above classification widely accepted. *However, there is little evidence for a common ancestor for both classes and no explanation about the evolutionary class separation.*

All class I aaRSs contain an insertion between the two halves of the ATP binding Rossmann fold. This insertion is called "Connecting Peptide 1" (CP1) (Starzyk, Burbaum *et al.*, 1989). In the smallest class Ic aaRS, tryptophanyl-tRNA synthetase (TrpRS), this insertion has 76 residues. In four of the large class Ia aaRSs (LeuRS, MetRS, IleRS, ValRS), the corresponding internal insertion has grown bigger in size into a more structured domain harboring an editing function that is specific for the most frequently mistaken amino acid (Jakubowski and Goldman, 1992). Similarly, the carboxyl-terminal anticodon binding domain might have evolved from a shorter primitive ancestral peptide as a later addition. Rodin and Ohno have proposed a unique hypothesis to explain the emergence of contemporary aaRSs based on sequence alignments. They found that the coding sequences for the HIGH and KMSKS class I signatures are very nearly antisense to the corresponding coding sequences of the class I motifs 2 and 1, respectively (Rodin and Ohno, 1995). They noted that when all of the intervening insertions are removed, the KMSKS and HIGH loops from class I aaRSs can be aligned approximately antiparallel with motifs 1 and

2 from class II aaRSs as sense and antisense strands of one gene. The earliest aaRS would have been active enzymes; otherwise they could not have participated in natural selection. The Rodin/Ohno observation therefore suggests directly that a constructed gene lacking CP1 should be enzymatically active. However, no experiment has been done to test this hypothesis.

In this chapter, an experiment that characterizes a redesigned, catalytically active TrpRS Urzyme consistent with Rodin and Ohno's observation will be discussed as the first experimental evidence supporting Rodin-Ohno hypothesis. Using computational algorithms, a 130-residue long Urzyme has been constructed by removing the internal 76 residues connecting peptide and truncating the carboxyl-terminal domain. This Urzyme shows significant activity in catalyzing amino acid activation compared to the native enzyme.

Materials and Methods

Linker Design to Replace Connecting Peptide 1

TrpRS Urzyme was constructed *in silico* by fusing two halves of the Rossmann fold. The connecting peptide 1, a 74-residue long fragment between T46 and G121, and the last 124 amino acids at the carboxyl-terminal domain were removed. The resulting 130-residue long sequence was submitted to Rosetta Design (Dantas *et al*., 2003; Kuhlman and Baker, 2004; Schueler-Furman *et al*., 2005; Dantas *et al*., 2006) with six linkers (A, AA, AGA, AAAA, AAGAA, AGAAGA) that vary in length and consist of short side chain amino acids alanine (A) and glycine (G). A zero length linker was also included. Up to 4 residues adjacent to each side of CP1 removal were allowed to mutate to facilitate restoration of molecular continuity. For each linker, sixty backbones were pulled out of the biological database in the Protein Data Bank (PDB) with assigned energy scores as the indicator for stability measurement.

Solvent Accessible Surface Area (SASA) Analysis to Improve Solubility

The solvent accessible surface area was compared between TrpRS Urzyme and the native full-sized enzyme in three different conformations, the open unliganded state, the transition state and the closed liganded product state, using Areaimol (CCP4, 1991). A threshold value of +50 $Å^2$ was set as limit for the exposure of hydrophobic residues caused by CP1 and ABD removal to identify the residues that would be subjected to mutation by Rosetta Design. The predicted mutations were introduced to TrpRS Urzyme in addition to the previously selected in the CP1 loop closure. Genes for two TrpRS Urzymes amino acid sequences either with or without these additional mutations designated redesigned (DES) and wild type (WT), respectively, were synthesized by Genscipt and optimized to eliminate rare codons.

Bacterial Strains and Plasmid Construction

Restriction enzymes and other molecular biology reagents were obtained from New England Biolabs and were used according to the manufacturer's protocols. Forward and reverse primers for PCR amplification were designed to be complementary to the Urzyme sequence, containing extra nucleotides coding for N-terminal FLAG and C-terminal 6xHIS tags. Desired restriction sites were also introduced in the primer design step. Following digestion with the appropriate enzymes, the PCR products were cloned into *E.coli* expression plasmid pET-42*a* for use in bacterial strain BL21(DE3)pLysS (Novagen). The D146A mutant in the DES fragment was prepared using the Invitrogen GeneTailor Site-Directed Mutagenesis System. All constructs were confirmed by DNA sequencing.

Protein Expression

All TrpRS fragments were expressed directly after transforming with plasmid DNA. After transformation, 100 μl competent cells were incubated in 1ml LB media at 37°C for 1 hr with shaking and then used to inoculate 100 ml growth media (LB plus 2% glucose, 30 μg/ml kanamycin, 34 μg/ml chloramphenicol). The inoculum was shaken at 250 rpm overnight at 37°C and then transferred to 1 l fresh LB with 30 μg/ml kanamycin and 34 μg/ml chloramphenicol. At OD_{600} = 0.4–0.8, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 1 mM final concentration to induce gene expression. Harvest began 14 hr after induction.

Purification of Fragments from Supernatant and Inclusion Bodies

Cell-free extract was prepared from frozen cell pellets by lysing cells in 50 mM Tris, 10% sucrose (pH 7.5), using a continuous flow pressure cell (Avestin), followed by centrifugation at 20,000 rpm for 30 min at 4°C, and the pellets were re-extracted in the same manner. Supernatants were combined and centrifuged at 27,500 rpm for 1 hr to yield the soluble fraction. The pellet (inclusion bodies) from this step and the

combined supernatant were used as starting materials for purification. The soluble fraction was then dialyzed against anti-FLAG washing buffer (1× PBS, 0.05% Tween-20, 5 mM ATP, 50 μM tryptophan, 10% glycerol, 50 mM KCl, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 4 mM $MgCl_2$ [pH 7.8]).

Renaturation from Inclusion Bodies

Optimized renaturation conditions were determined from an incomplete factorial design (Chester *et al*., 2004; Yin and Carter, 1996) of nine different sets of conditions in which the temperature, protein concentration, KCl concentration, pH, dialysis protocol, and presence of substrates were varied randomly and rebalanced. Inclusion bodies were first dissolved and diluted in solubilization buffer (20 mM HEPES [pH 7.8], 6 M urea, 50 mM or 250 mM KCl, 5 mM $MgCl_2$, 1 mM EDTA, 1 mM PMSF, 1 mM β-mercaptoethanol, 5% glycerol [Campbell and Carpenter, 1995] with and without substrates) to OD_{280} = 2–10 (approximate total protein concentration of 2–10 mg/ml). Supernatants after further centrifugation (17,000 rpm for 30 min at 4°C) were dialyzed three times against ten volumes of the same buffer at 4° or 37°C, without urea. Before the final dialysis, bags were opened and 10 μl of anti-FLAG beads (Sigma) were added per ml of protein. FLAG-bound material was eluted in the same solution with 50 μg/ml of FLAG peptide solution. Multiple regression of specific activities against the experimental variables identified three factors (low protein concentration, high pH, and low KCl concentration) responsible for (R^2 = 0.87) of the variation in the data with t test probabilities of $\sim 10^4$.

Purification on Anti-FLAG Resin Beads

Prewashed anti-FLAG bead suspension was added (10 μl/ml) to supernatants, which were shaken overnight at 4°C. Fragments bound to the beads (recovered in Bio-Rad Hi-Trap Q columns) were eluted with FLAG peptide (0.5 μl/ml).

Quantitation of Specific Activity

Both immunoblotting with anti-FLAG antibody and silver staining were used to construct standard curves for estimating protein concentrations. As the silver-stained gels showed that the samples were quite pure, we determined most concentrations by using silver staining.

^{32}PPi Exchange Assays for Tryptophan Activation

Fragment and empty-vector eluates from anti-FLAG beads were concentrated 8- to 25-fold before being assayed. Samples were assayed using the traditional ^{32}P-pyrophosphate exchange as described (Joseph and Muench, 1971) except that a 3-fold reduction of background ^{32}P counts was achieved by collecting the charcoal containing labeled ATP on disposable spin columns, washing, and eluting the bound ATP with 50 μl of pyridine at 37°C.

Results

Computational Phase to Design a Linker to Ensure TrpRS Urzyme Molecular Continuity

Based on the multiple sequence alignments (MSA) of known class I and class II aaRSs, a gene encoding the highly conserved residues in two disjoint TrpRS Rossmann fold fragments has been constructed (Pham *et al*., 2007). In this 130-amino acid Urzyme, the connecting peptide (CP1) is removed and the C-terminal domain is truncated, thereby exposing the molecule to two anticipated problems: structural disruption at the site of CP1 removal and, more severely, instability and/or insolubility due to the loss of two thirds of its stable configuration. Using the Rosetta algorithm, a short linker of varying length, from 0 to 6 residues, containing short side chain amino acids alanine and glycine, has been scanned to ensure molecular continuity. Furthermore, three residues adjacent to the sites of T46 and G121 were allowed to mutate (backbone melting), facilitating the joining of two free ends. The Rosetta Design program assigned lowest energy score for the direct linkage between threonine and glycine, indicating that a stable construct could be possible without inserting any residue into the gap created by removing CP1 (Figure 6). This construct had five mutations, created by allowing relaxed backbones, illustrated by the blue spheres at the site of removal (Figure 7A).

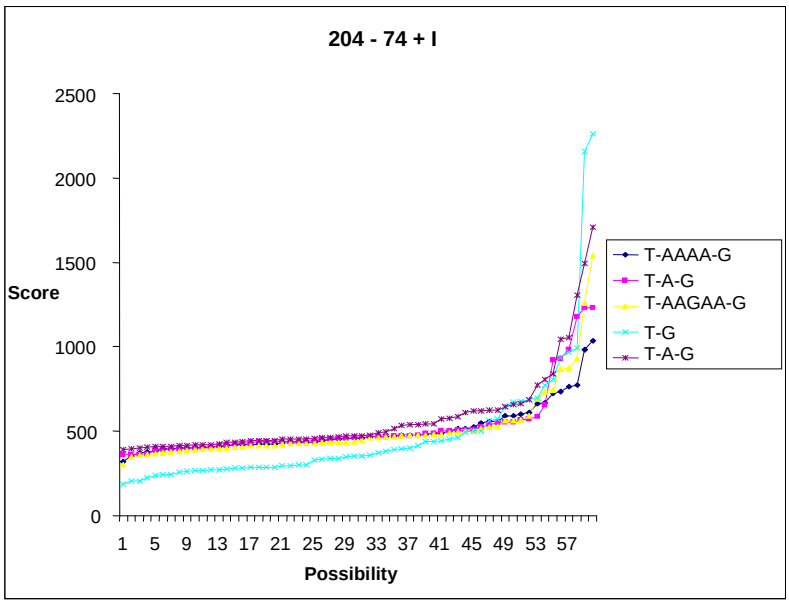


Figure 6: The Rosetta energy profile shows a decisively lower energy score for most constructs of 130-residue long after CP1 and CTD truncation. The lowest energy of these constructs, with zero length linker (direct linkage between T46 and G121) and five relaxed side chains adjacent to the site of connecting peptide 1 removal (Pham, unpublished) constituted the wild type Urzyme.

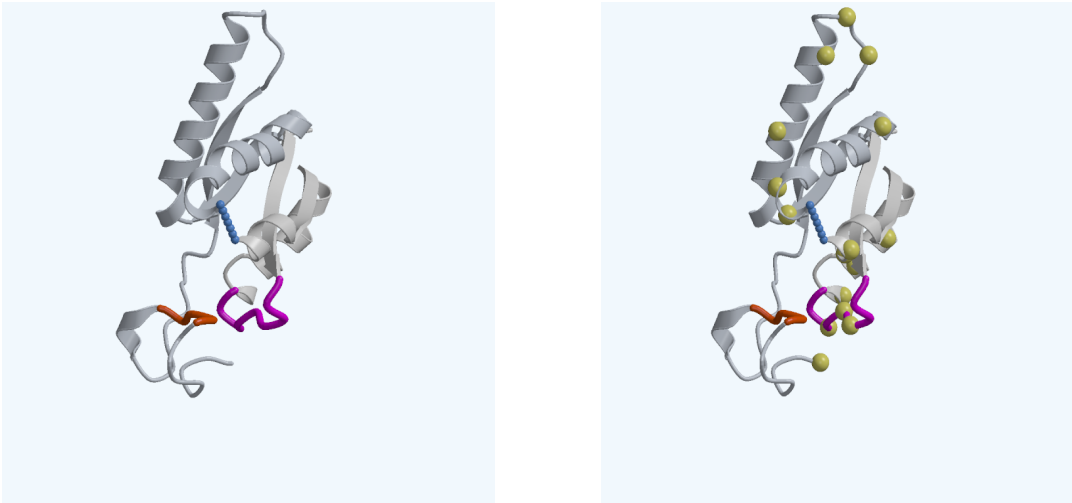
Solubilizing and Stabilizing the Wild Type TrpRS Urzyme

To address expected instability and insolubility problems, the profile of the accessible surface area (ASA) for each residue in the computationally simulated Urzyme has been compared with the native enzyme using Areaimol (CCP4, 1991). From 60 positively increased ASA values, indicating changes in exposure due to the removal of CP1 and the ABD, the 19 highest were selected and submitted to Rosetta for mutation suggestion (Table 1). It should be noted that, excepting Tyr91, none of these mutations involved highly conserved residues (Table 2).

Amino acid	Residue	ΔS(PROD)	ΔS(PreTS)	ΔS(Open)
VAL	13	9	99	19
ILE	14	85	71	102
ILE	16	57	62	58
TYR	19	98	74	83
LEU	23	57	56	63
TYR	36	90	86	65
PHE	37	84	88	95
ILE	39	124	127	124
LEU	122	132	120	89
LEU	123	115	119	117
LEU	135	130	119	112
TYR	136	87	81	91
LEU	153	64	58	56
LEU	157	52	55	51
PHE	161	72	75	76
TYR	165	89	80	74
LEU	168	75	76	69
LEU	186	105	91	109
ILE	204	114	119	134

Table 1: Comparison of solvent accessible surface area between WT Urzyme and full length TrpRS in three different configurations. PROD: product-bound, Pre-TS: pre transition state, Open: unliganded (Carter, unpublished).

The second Urzyme, called the redesigned (DES), harboring 12 mutations predicted by Rosetta to increase solubility and stability is the result of the final



A **B**

Figure 7: Two TrpRS Urzyme constructs after computational design. A: wild type Urzyme (WT) retains native sequence plus five mutated residues at the site of CP1 removal. B: redesigned Urzyme (DES) harboring twelve mutations (see Table 2) predicted to improve stability and solubility (adapted from Pham et al., 2007).

computational processing step (Figure 7B). This redesigned Urzyme has a total number of 17 mutations compared to the native sequence (Table 2).

Residue (WT resnum)	Wild Type	Designed (Des)	ΔG_{stat}
13 (13)	Val	Met	0.13
14 (14)	Ile	Leu	0.40
16 (16)	Ile	Pro	0.42
19 (19)	Tyr	Trp	0.43
23 (23)	**Leu**	**Gly**	0.16
39 (30)	Ile	Met	0.31
44 (44)	*Ala*	*Tyr*	*0.18*
45 (45)	*Ile*	*Pro*	*0.15*
46 (46)	*Thr*	*Val*	*0.17*
47 (122)	***Leu***	***Glu***	*0.12*
48 (123)	***Leu***	***Thr***	*0.32*
61 (135)	Leu	Gly	0.16
79 (153)	**Leu**	**Lys**	**0.19**
87 (161)	**Phe**	**Tyr**	0.24
91 (165)	**Tyr**	**His**	**0.92**
74 (168)	Leu	Phe	0.13
130 (204)	Ile	Leu	0.41

Table 2: List of all mutations introduced to the TrpRS Urzymes including five residues which were changed from hydrophobic to hydrophilic in the wild type (WT) and redesigned (DES) Urzymes, respectively (bold). Free energy differences () were also calculated showing minimal effect of the mutations (Pham et al., 2007). A ΔG_{stat} of less than 0.45 is a measure of relative sequence conservation as defined by Lockless *et al*., 1999.

Expression and Purification of a Potentially-toxic Urzyme

The target sequence for WT and DES TrpRS Urzymes were synthesized and delivered in pUC57 by Genscript. This expression system is known for a high copy number, between 150 and 200, which would result in higher expression level of the protein of interest. However, the Urzyme is likely to be toxic for the cells due to the predicted low specificity. Thus, a lower copy number plasmid system was desired. pET42, a system derived from pUC57, was chosen for expression. With 15-20 copies of the plasmid per cell, the protein of interest would be less highly expressed which might affect the total yield but increase the Urzyme stability.

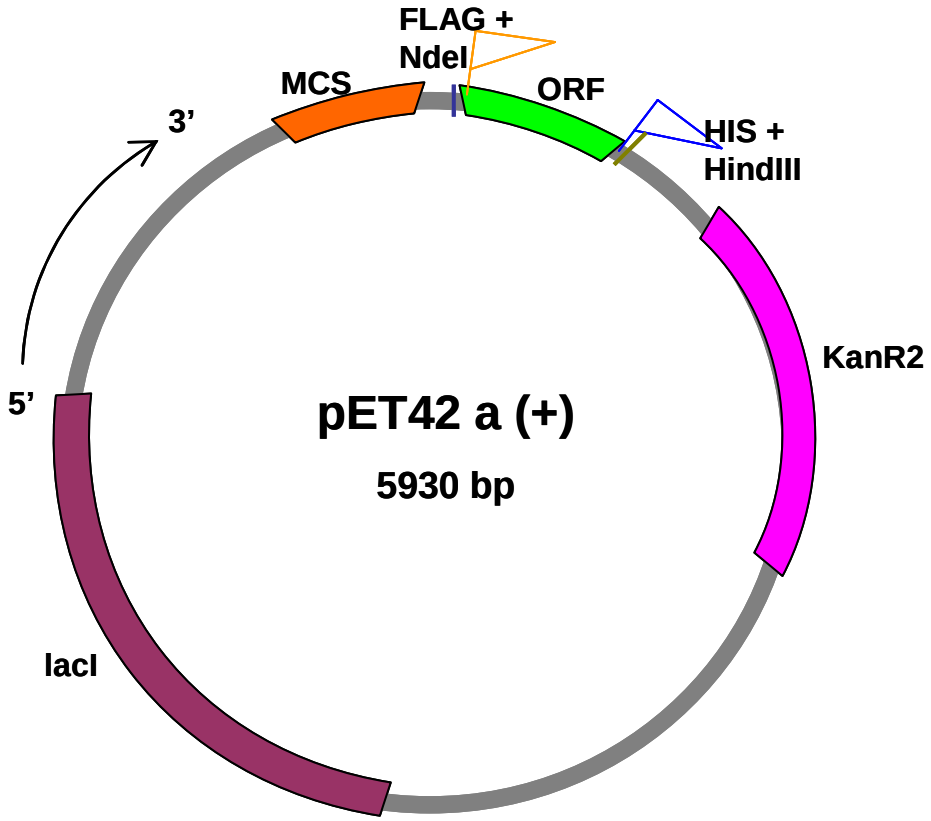


Figure 8: pET 42 a (+) expression vector showing relative positions of elements introduced. The open reading frame (ORF) of the TrpRS Urzyme genes was tagged with an N-terminal FLAG and a C-terminal 6xHIS (Pham, unpublished).

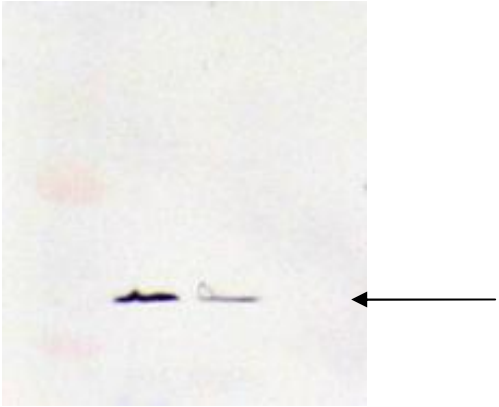
Two tags were introduced to facilitate detection and purification of the enzyme. The final expression system is shown in the Figure 8. pET 42a (+) containing the target gene is transformed into BL21(DE3) pLysS; this expression system has lower basal expression due to inhibition of T7 RNA polymerase by lysozyme, minimizing potential toxicity of the desired product prior to induction.

TrpRS Urzyme Expressed in the pET42a System Was Insoluble

Using an incomplete factorial design, an optimized expression condition was set for the TrpRS Urzymes (Table 3, blue). As showed in Figure 14 below, the amount of soluble Urzyme could not be detected with immunoblotting which is more sensitive than Coomassie Blue or silver staining methods. This indicated that the DES TrpRS Urzyme, after designed modification, was still unstable in the supernatant fraction.

Sample	Temperature (^{0}C)	EtOH (%v)	Induction time (hr)	Pellet weight (g)	Band intensity
1	15	0	4	0.27	5
2	15	0.064	6.7	0.26	0
3	15	0.04	12.1	0.25	4
4	15	0	20	0.28	6
5	15	0.08	20	0.24	0
6	25	0.08	4	0.24	0
7	25	0.013	10.4	0.38	7
8	25	0.044	15.9	0.27	3
9	37	0.034	4	0.33	8
10	37	0.08	11.5	0.23	2
11	**37**	**0**	**14.1**	**0.43**	**10**
12	37	0.045	20	0.23	3

Table 3: Expression condition optimization for the TrpRS Urzymes. The 11th sample (blue) had highest yield measured.



M T P S

Figure 9: **Fractionation of TrpRS Urzyme in the supernatant (S) and the inclusion bodies pellet (P) compared with total yield (T).** The Urzyme was immunoblotted with anti-FLAG antibodies and showed as the expected molecular weight (arrow position) (Pham, unpublished data).

Renaturation was Efficient to Solubilize the Redesigned, but Not the Wild Type Urzyme

Since most of the Urzyme was in the inclusion bodies, a refolding method using urea to dissolve the pellet was chosen to recover the fragment in soluble form. Using the same incomplete factorial design method, 9 different combinations of protein concentration, dialyzing buffer pH, salt concentration and renaturation temperature were designed to select an optimized set of conditions for both WT and DES Urzymes (Table 4, bold). However, while being dialyzed, the wild type TrpRS Urzyme showed signs of aggregation which led to the decision to further characterize only the redesigned construct.

Sample	Protein concentration (OD_{280})	pH of dialyzing buffer	[KCl] (mM)	Renaturation temp (^{o}C)	Relative specific activity
1	2	8.6	50	25	8
2	10	7.0	50	4	1
3	6	5.4	250	4	2
4	6	7.0	250	4	2
5	10	5.4	50	25	2
6	6	8.6	50	25	7
7	**2**	**8.6**	**50**	**4**	**9**
8	2	5.4	250	25	7
9	2	7.0	250	4	3

Table 4: Renaturation condition optimization for the TrpRS Urzymes. The 7th sample (blue) had highest relative specific activity.

Pyrophosphate Exchange Assay Show Significant Activity for the Renatured TrpRS Urzyme

The renatured Urzyme amino acid activation activity, as showed in the below table, is 10^5 times weaker than that of the native enzyme. In the table, the native TrpRS activity is normalized to 1 unit. It is interesting that the wild type Urzyme, besides being more insoluble, also has lower activity. This indicates that the

redesigned Urzyme, a minimal catalytic domain (MCD), might have had more proper folding thus becoming more stable, soluble, and active.

An active site mutant, D146A, showed 20 fold higher activity compared to the redesigned Urzyme. The same mutant introduced in the full length enzyme reduces the activity 200 fold. This suggests a different catalytic mechanism for the Urzyme. More importantly, it proves that the activity detected is coming from the Urzyme itself since a targeted mutation allows its modulation.

Protein	Specific activity (Prod/mol/s) max; ave ± stdev (replicates)	Relative specific activity
Wild type TrpRS	64888	1
D146A full length	326, 308±25 (2)	0.005
Des MCD	1.36; 1.26±0.16 (4)	0.00002
Wild type MCD		
D146A Des MCD	26.5; 20.2±8.9 (2)	0.0004
	k_{cat}/Km (Prod/mol/s)	
Ribozyme (*40*)	0.38	.000006

Table 5: Activity of the refolded TrpRS Urzymes – wild type (WT) and redesigned (DES) compared with the full length enzyme, native D146A mutant and ribozyme (Pham et al., 2007).

For reference, the activity of an extensively modified ribozyme (Kumar and Yarus, 2001) was ~3 times slower than the redesigned Urzyme (Des MCD).

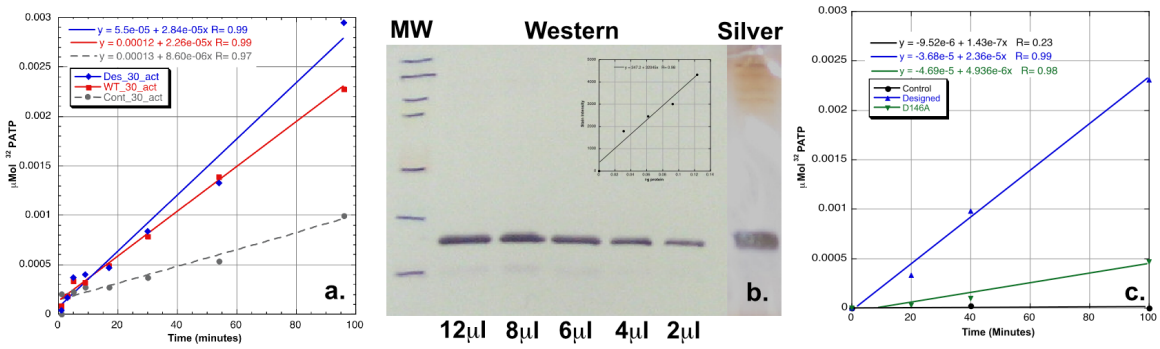


Figure 10: The TrpRS Urzymes, quantitated by immunoblotting and silver staining (b.), were used for pyrophosphate exchange assay. The noise level in (a., dotted green line) was subtracted as background to produce (c.) (Pham et al., 2007).

Discussion

Rosetta Design and Choice of Mutations

As indicated in the Materials and Methods section, the number of biological sequences scanned was limited to sixty for each linker. The TrpRS Urzymes synthesized have the lowest energy scores within the limit, however, might not be the best among all naturally occurring sequences in the database. It is notable that in all seven linkers ranging from 0 to 6 amino acids long, the zero length linker is shown to be most stable. This result might imply that the 5 Ångstrom gap between residues threonine 46 and glycine 121 could have once not existed, and thus, that the connecting peptide 1 would be a later insertion. The computational construction phase, however, only tries to design linkers using two short side chain amino acids, alanine and glycine, in a certain number (6) of all possible combinations. The rationale of using A and G is to avoid any interaction between the inserted side chain and the main backbone. Nevertheless, this brings more constraints to the optimization process. Another constraint is the number of residues adjacent to the side of T46 and G121 that are allowed to mutate. The Urzyme design, indicated to be stable in a 40 ns molecular dynamics experiment (H. Hu, data not showed), is still open for more improvement.

Of all mutations predicted to stabilize and solubilize the TrpRS Urzyme, only five are from hydrophobic to hydrophilic (Table 3), which would directly change the interaction with solvent. It has been noticed that Rosetta Design has a tendency to place large patches of hydrophobic residues together (Liu and Kuhlman, 2006). Since a significant part of the surface area of the full length TrpRS has been removed, exposing the hydrophobic core, the introduction of more hydrophobic residues at different locations in the Urzyme might help recover the proper folding. However, the

drawback of this is that the designed molecule is more likely to aggregate. A more stable and soluble Urzyme could be the next target as the algorithm is being rapidly developed.

Assaying Weak Amino Acid Activation Activity after Renaturation and tRNA Acylation Activity

The plasmid carrying the TrpRS Urzyme was transformed into BL21 cells resulting in the over-expression of the desired protein. Over-expressed protein usually accounts for approximately 50% of the total protein. Since possessing a 10^5-fold weaker amino acid activation activity, the Urzyme may be toxic for the cells. Evidence that this is the case includes the fact that the plasmid is rapidly lost, even when selective pressure is maintained by using antibiotics. The presence of the majority of the Urzyme in the inclusion bodies is an indicator for aggregation. As previously predicted, the Urzyme activity would be much weaker, thus, could be easily interfered by the full length enzyme activity brought by the host cell. Since this contaminating activity comes from the soluble form that stays in the supernatant, purifying the molecule from the insoluble fraction does have an advantage in eliminating the host cell contaminating activity. However, the wild type Urzyme tends to go back to an aggregated form in the renaturation process. This phenomenon led to the decision to do further characterization extensively with the redesigned construct. Together with a molecular dynamics simulation done by a colleague, the higher tendency of redesigned Urzyme to stay in soluble fraction compared to the wild type confirms the success of Rosetta Design to stabilize and solubilize the Urzyme. Nevertheless, no structural analysis has been done beyond molecular dynamics to examine whether the refolding process is complete and accurate. A soluble form of TrpRS Urzyme is still preferable. It is important to notice the D146A active site

mutant, which reduces the reaction rate of the native enzyme by a factor of 200, surprisingly increases that of the redesigned Urzyme by a factor of 20. This result was unexpected, however, it indicates that the detected weak activity is not an artifact, and possibly, implies a different mechanism for the Urzyme.

Since amino acid activation is the first half of the reaction that is not dependent on tRNA, a second assay for cognate tRNA acylation as described by Wolfson and Uhlenbeck is necessary (Wolfson and Uhlenbeck, 2002). TrpRS Urzyme lacks the connecting peptide 1 containing the tRNA interacting helix but still retains other elements thought to be part of the tRNA binding site (i.e. residues 142-150 of the αE helix) that could assist the second half of the aminoacylating reaction (Jia *et al*., 2002). It will also be interesting to co-express the dissected domains, CP1 and CTD either *in cis* or *in trans* to unravel the role of those in catalysis.

Implications for TrpRS Evolution

The exclusive differences in structure and catalytic mechanism (e.g. tRNA acceptor stem binding mode, hydroxyl group specificity) described earlier strongly supports the idea of two separate ancestors for class I and class II aaRS. At first glance, the hypothesis proposed by Rodin and Ohno that the two classes might have originated from two complementary strands of one gene is very unusual. If that is the case, the deleterious (e.g. lethal or nonsense) mutations occur in one strand may lead to inactivation of the gene on the other strand. However, one might want both aaRS classes encoded in one DNA fragment because that would save the genetic material and more importantly, allow synchronized expression, especially in a pre-cellular environment. One of the very first pieces of evidence for the Rodin-Ohno hypothesis came from structural alignment analysis. The DNA encoding one class's conserved motifs is used to generate a complementary strand, which is then computationally

decoded into amino acid sequence. It is interesting that the resulting peptide showed non-random homology to the signature sequences of the other class (Rodin and Ohno, 1995).

Another important clue is the existence of additional functional domains among aaRSs. The two variable-length connecting peptides lying between two segments of the conserved Rossmann fold were once thought unrelated to the catalytic activity. However, the connecting peptide 1 (CP1) is now determined to be responsible for the proofreading (editing) activity of at least five aaRS. Connecting peptide 2 (CP2) in *E. coli* LeuRS, whose sequence is very similar to *E. coli* leucine-specific binding protein (Williamson and Oxender, 1990), is considered as a later insertion bringing more fidelity to the aminoacylation reaction (Zhou *et al*., 2008).

In an article published in 2002, Carter and Duax have pointed out the homologies between a real sense/antisense coding system and the two aaRS classes. Class I and class II aaRS structurally resemble *Achlya klebsiana* glutamate dehydrogenase (GDH) and HSP70-like chaperonin (a heat shock protein), respectively (Carter and Duax, 2002). The two latter proteins are shown to result from complementary strands of the same contemporary gene (LéJohn *et al*., 1994). This interesting observation has gathered controversial feedback. Williams *et al.* (2009) has recently argued about the accuracy of the methods used by the authors in order to deduce their conclusion. The TrpRS Urzyme model initiated in this chapter suggests a possibility that two classes were once encoded by one ancestral gene, and that later insertions were added to enhance the overall performance of aminoacylation. Hence, this is important evidence implicating aaRS sense/antisense origin, reinforcing Rodin-Ohno hypothesis (Rodin, Rodin and Carter, 2009).

Conclusion

Solving the question of how two contemporary class I and class II aaRS might have emerged and evolved would shed light on the involvement of these enzymes in current pathways and probably allow the prediction of new functions. This chapter describes a TrpRS Urzyme constructed by fusing two discrete halves of the ATP binding Rossmann fold that shows significant amino acid activation activity. The 130-residue long peptide, lacking the connecting peptide 1 and the carboxyl-terminal anticodon binding domain that make up 60% the mass of the active dimer, activates tryptophan with 10^5 times weaker activity. However, compared to non-catalyzed reaction on a logarithmic scale, it accelerates the rate by a factor of 10^9, and thus lowers the activation free energy by –12.4 Kcal/mol. The overall rate is accelerated by 10^{14}, and therefore lowers the activation energy by –19.24 Kcal/mol. The ratio of the Urzyme's catalytic rate enhancement to that of the native enzyme in terms of activation free energy is therefore –12.4/-19.24 = 0.644 ~65%, and thus, is a very significant fraction of the overall rate enhancement of the native enzyme. It is important to know that this is not contaminating activity from the host cell TrpRS since the empty vector control has been substracted as background noise. Besides, TrpRS Urzyme has been purified through renaturation from the inclusion bodies, separately from the soluble fraction that could contain host-cell native form of the enzyme. Further experiments are to be done to characterize the Urzyme to understand the modular complementation of the removed domains. More importantly, this Urzyme might be a potential candidate for the ancestral class I aminoacyl-tRNA synthetase postulated in Rodin-Ohno hypothesis. The procedures described in this chapter can serve as a paradigm to construct a class II Urzyme to further verify the sense/antisense relationship hypothesized to be once existed.

Acknowledgment

The computational design phase received major support from Rosetta Design developing team in the Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill. I would like to thank Dr. Kuhlman and G. Butterfoss for their contribution to this initial step of the project, L. Li and V. Weinreb, our lab members, for their assistance in molecular manipulations.

This work is supported by NIH grant GM078227-01 and is published as Pham *et al*. (2007). A minimal TrpRS catalytic domain supports sense/antisense ancestry of class I and II aminoacyl-tRNA synthetases. *Molecular Cell* **25(6),** 851-62.

CHAPTER 3

DETERMINATION OF AN EXPRESSION SYSTEM TO MAXIMIZE THE YIELD OF TRPRS URZYME IN THE SOLUBLE FRACTION

Abstract

The next step to verify the accuracy of protein design is expression and characterization of the expressed construct. As previously described, a redesigned tryptophanyl-tRNA synthetase Urzyme showed significant tryptophan activation activity, and when refolded, stayed in the soluble form, unlike the wild type which had a pronounced tendency to aggregate. However, the majority of the Urzyme was in the inclusion bodies and required renaturation, which might lead to improper folding. Using different expression systems, one often can bring the protein of interest to the soluble fraction with a highly soluble carrier protein. This chapter is devoted to determination of a suitable vector to solubilize the redesigned TrpRS Urzyme. Of the two systems described, pET SUMO (Small Ubiquitin-related Modifiers) (Novagen) increased the total yield but not the solubility. The other vector, pMAL-c2x encoding the Maltose Binding Protein (MBP) (New England Biolabs), improved the soluble yield substantially. More importantly, when digested with Factor Xa to remove the carrier protein MBP, the Urzyme was stable and showed tryptophan activation activity comparable to that of the original refolded Urzyme. Michaelis-Menten kinetics of this solubilized TrpRS Urzyme are described in Chapter 4. Further characterization remains a future direction.

Introduction

Protein expression and purification are essential for further manipulations in biochemical research. The need for having large amounts of protein synthesized and purified has led to the development of different expression systems, efficient purification kits and a large database for yield, purity optimization methods as well as troubleshooting tips. It has been thoroughly discussed in the book of Current Protocols in Molecular Biology (Wiley Interscience, 1991) about the most widely used expression systems. Among those, *Escherichia coli* has become an extremely popular host with a certain number of advantages compared to the others.

E. coli expression systems can be controlled at both transcriptional and translational levels. With different mutations at the origin of replication, one is capable of manipulating the speed of bacterial growth and thus, expression. It has been suggested that low plasmid copy number would be appropriate for desired proteins that are potentially toxic for the cell. For that purpose, pET, with a copy number ranging from 15 to 20 per cell, is appropriate. As discussed in the previous chapter, the TrpRS Urzymes have been subcloned into pET42a (+) vector, which would solve the aggregation problem of the Urzyme, suggested by the fact that the plasmids are lost quickly, in contrast to the wild-type TrpRS-coding plasmid. Further expression in a tightly controlled system, BL21 pLysS provides additional protection against toxicity.

Once efficiently expressed, a protein can be in either soluble form in the supernatant fraction or denatured state in pellet fraction that includes the inclusion bodies. Purifying proteins from inclusion bodies through renaturation after an initial denaturing step to solubilize the pellet fraction by urea or guanidium hydrochloride has been described by many groups. However, additional complicated procedures

have become necessary to assure proper folding of the protein of interest after renaturation. Attention to folding is necessary especially if the protein has a marginal stable structure and catalytic activity. This concern gives rise to methods of using a highly soluble tag, conjugated to the molecule as fusions. A variety of carrier proteins with different sizes are available for consideration, including maltose binding protein (MBP), glutathione-S-transferase (GST), and small ubiquitin-related modifiers (SUMO). Beside the effect of shifting the expression of the previously insoluble desired protein into the soluble fraction, one might benefit from using the carrier for further purification steps. Resins conjugated with substances that have high affinity for carrier proteins have become commercially available, as well as the necessary proteases to separate the target from its carrier.

The phenomenon described with the wild type TrpRS Urzyme in the refolding process (Chapter 2) implies its low solubility and high tendency to aggregate. Besides paying necessary routine attention to internal cleavage sites by the recommended proteases used to digest the carrier-ORF linker, there is a possibility that the Urzymes may aggregate and go back to the insoluble fraction without the presence of the carrier. This will lead to the potential loss of activity, hence, requires special considerations when selecting a fusion system.

Maltose binding protein fusion is synthesized from the pMAL plasmid (Figure 11). New England Biolabs has two main sub-systems to express the protein of interest either in the cytoplasm or periplasm. The depicted pMAL-c2 plasmid belongs to the former sub-system, which would utilize Factor Xa to cleave off the maltose binding protein moiety. The protease's recognition site is Ile-Glu-Gly-Arg (IEGR).

Figure 12 is an illustration of glutathione-S transferase fusion system developed by GE Healthcare and Life Sciences. The pGEX plasmid contains a cleavage site for

the Tobacco Etch Virus (TEV) protease, whose recognition sequence (ENLYFQG is the most commonly used, with some variations at non-critical positions) is longer than that of Factor Xa. As a result, this system is preferred for reducing non-specific internal cleavage.

SUMO is an expression system recently developed to serve the purpose of solubilizing proteins known to be difficult when it comes to expression in soluble fraction (Figure 13). The summary of most commonly used carrier proteins in Table 6 indicates that SUMO tag has average size compared to others. Membrane proteins, for example, have proved to be successfully expressed and purified in SUMO conjugated form (Marblestone et al., 2006). The authors also point out that SUMO protease is much more efficient than TEV protease. The former binds it substrate with a higher affinity and is a faster catalyst. Collectively, the SUMO system has become the first choice to solubilize and purify protein efficiently.

The TrpRS Urzyme has been improved for solubility and stability by Rosetta algorithms as indicated in the previous chapter. Molecular dynamics experiment have shown that the structure is stable; however, when expressed in the low copy number plasmid pET42, the wild type Urzyme aggregated in urea dialyzing buffer even in the presence of substrates (ATP and tryptophan). Since the predicted activity of TrpRS Urzymes is much weaker than the full length enzyme, a high yield is crucial for better detection of activity. However, use of many purification steps reduces the yield to the point where it is not satisfactory. Refolding poses another problem, which is the unknown efficiency of renaturation process. In other words, it would become necessary to determine what percentage of the Urzyme is properly refolded. To avoid further complication with the TrpRS Urzyme constructs, the open reading frame has

been subcloned into different expression vectors known to increase the fraction of the desired protein in stable and soluble form.

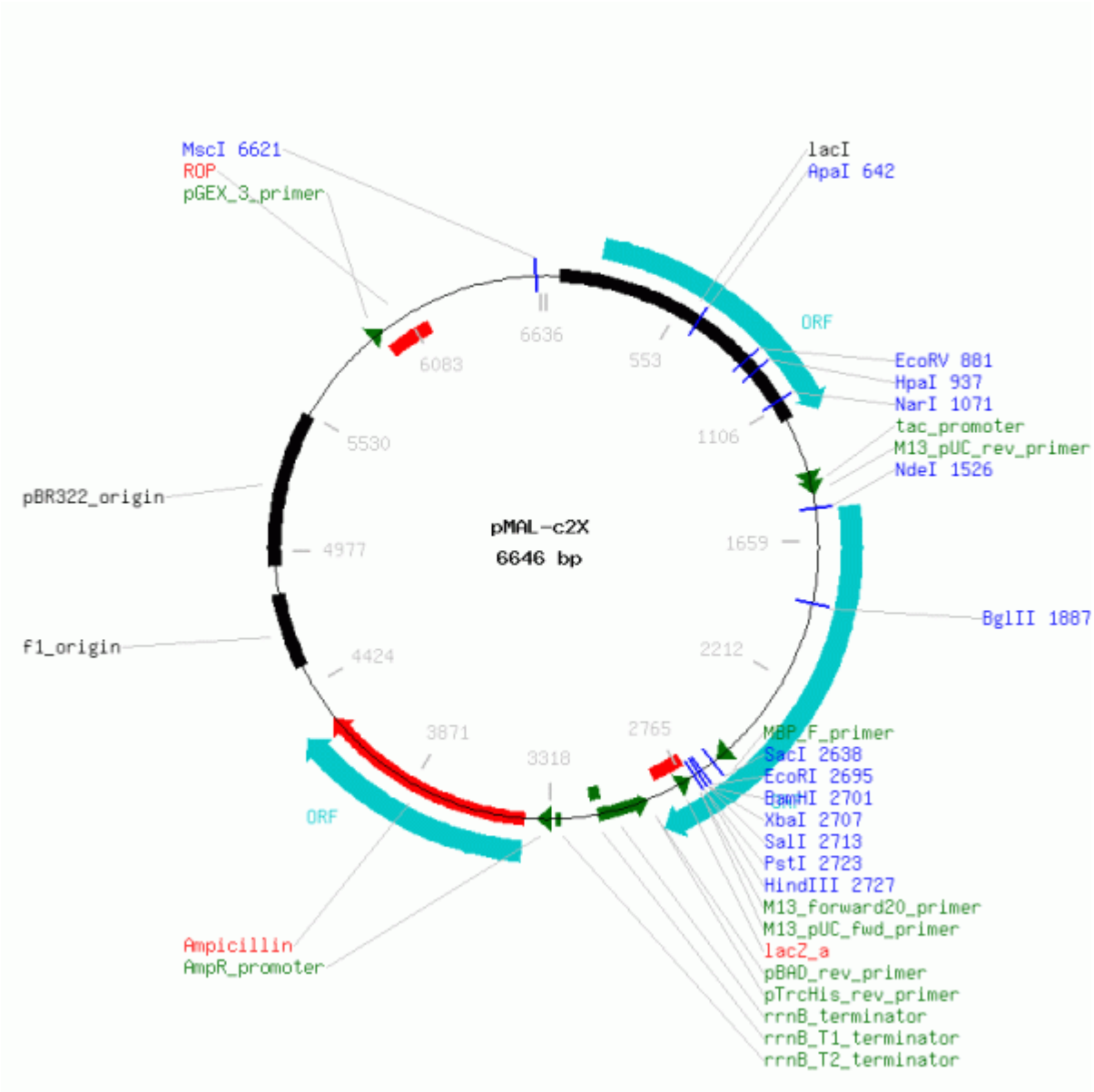


Figure 11: pMAL-c2x expression vector (New England Biolabs) uses the maltose binding protein (MBP or MAL) as the carrier protein.

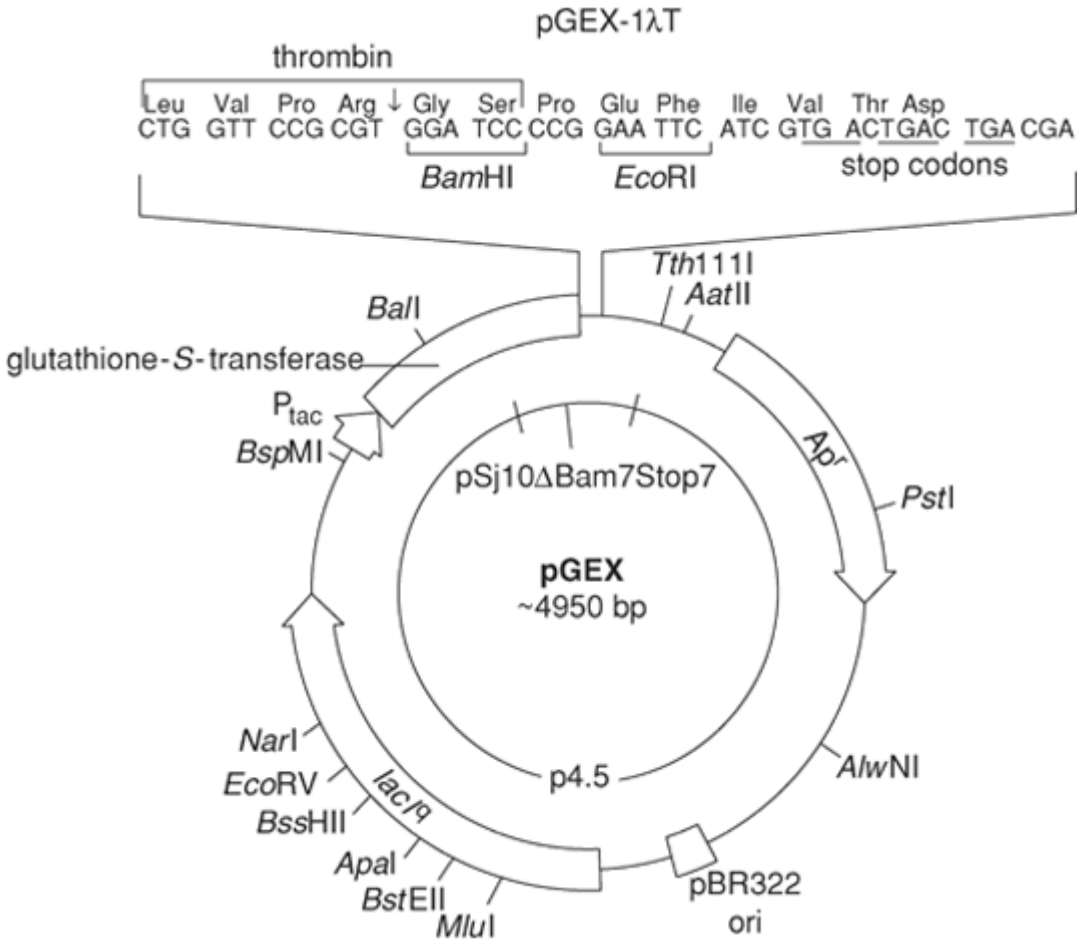


Figure 12: GST is one of the most popular systems used for protein expression and purification. The pGEX plasmid (GE Healthcare & Life Sciences) includes glutathione-S transferase gene whose product assists in affinity purification (Harper and Speicher, 2008).

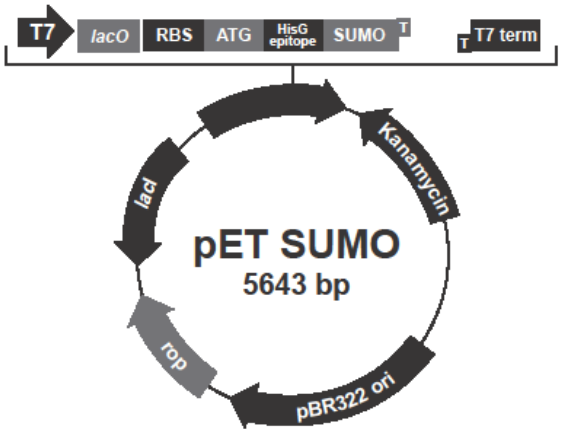


Figure 13: **pET SUMO system from Novagen showing the SUMO carrier protein at the N-terminal of the open reading frame** (Manufacturer's manual).

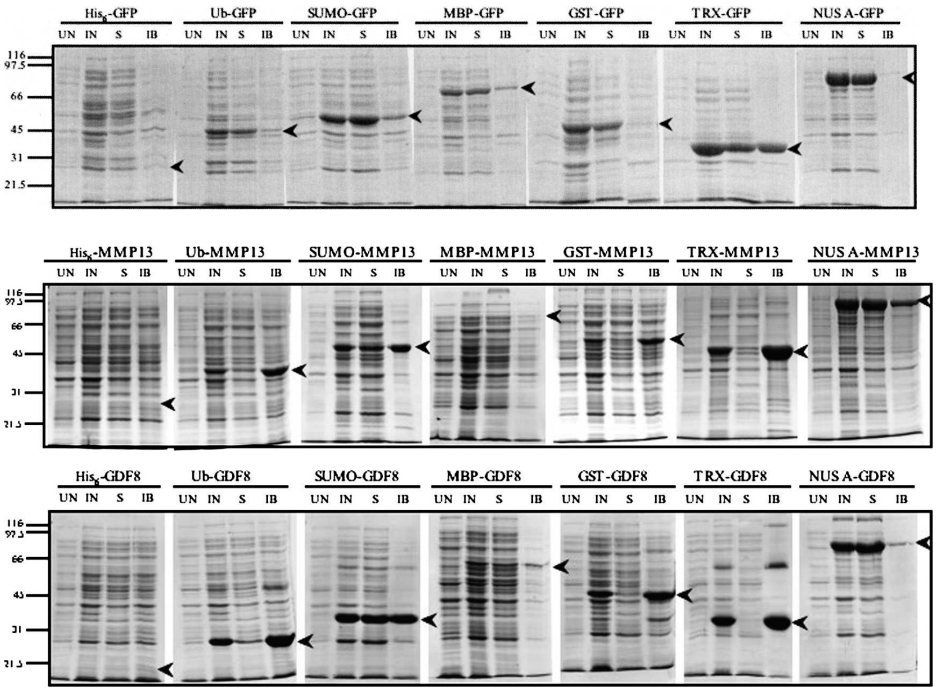


Figure 14: Comparison expression levels in soluble and insoluble fractions with different carrier proteins to solubilize difficult-to-express proteins (Marblestone *et al.*, 2006).

Tag	Residues	Cleavage site	Size (kDa)
Poly-His	6	—	0.84
Ubiquitin	76	ENLYFQ'GXX	8
SUMO	100	GG'XXX	11.5[a]
MaltoseBindingprotein	396	ENLYFQ'GXX	40
GlutathioneS-transferase	211	ENLYFQ'GXX	26
Thioredoxin	109	ENLYFQ'GXX	12
NUSA	495	ENLYFQ'GXX	55

[a] SUMO migrates aberrantly at 20kDa in an SDS–polyacrylamide gel.

Table 6: Carrier proteins frequently used to assist solubility and affinity purification (Marblestone *et al*., 2006).

Materials and Methods

Subcloning from an Original Vector to the Target Plasmid

Restriction enzymes and other molecular biology reagents were obtained from New England Biolabs and were used according to the manufacturer's protocols. Forward and reverse primers for PCR amplification were designed to be complementary to the Urzyme sequence, containing extra nucleotides coding for N-terminal FLAG tag and C-terminal 6xHIS tag. Desired restriction sites were also introduced in the primer design step. Following digestion with the appropriate restriction enzymes, the PCR products were cloned to *E.coli* expression plasmids pET SUMO (Novagen) or pMAL-c2x (NEB) for use in bacterial strain BL21(DE3)pLysS (Novagen). All constructs were confirmed by DNA sequencing.

Protein Expression

All TrpRS fragments were expressed directly after transforming with plasmid DNA. After transformation, 100 μl competent cells were incubated in 1ml LB media at 37°C for 1 hr with shaking and then used to inoculate 100 ml growth media (LB plus 2% glucose, appropriate antibiotics at 30 μg/ml for kanamycin, 34 μg/ml for chloramphenicol, 100 μg/ml for ampicillin). The inoculum was shaken at 250 rpm overnight at 37°C and then transferred to 1 l fresh LB with the same antibiotics concentrations. At OD_{600} = 0.4–0.8, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 0.3 or 1 mM final concentration to induce gene expression. Harvest began 3 or 14 hr after induction.

Analysis of Fragments from Supernatant and Inclusion Bodies

A cell-free extract was prepared from frozen cell pellets by lysing cells in 50 mM Tris, 10% sucrose (pH 7.5), using sonication in pulse mode (3x15 seconds) combined

with a continuous flow pressure cell (Avestin) when indicated, followed by centrifugation at 14,000 rpm for 30 min at 4°C.

Quantitation of Soluble Yield in the Supernatant Compared with Insoluble Fraction in the Inclusion Bodies

Both immunoblotting with anti-FLAG antibody and Coomassie Blue staining were used for estimating protein concentration. As the silver-stained gels showed that the samples were quite pure, we determined most concentrations by using Coomassie Blue staining combined with the absorbance obtained from a Nanodrop spectrophotometer.

Factor Xa Cleavage

The MBP fusion was digested with Factor Xa (NEB) overnight at 4°C with rotating. The ratio of substrate and enzyme is 1:50 (w/w). Concentrate fusion was diluted to a final concentration of 10% glycerol from stock stored in 50% glycerol (1:5 dilution). The digestion mixture was used directly for tryptophan activation assay.

^{32}PPi Exchange Assays for Tryptophan Activation

Samples were assayed using the traditional ^{32}P-pyrophosphate exchange as described (Joseph and Muench, 1971) except that a 3-fold reduction of background ^{32}P counts was achieved by collecting the charcoal containing labeled ATP on disposable spin columns, washing, and eluting the bound ATP with 50 μl of pyridine at 37°C.

Results

The SUMO Expression System Improved the Yield but Not the Solubility

As shown in Figure 14A, in the original pET42 a system, most of the protein was in the inclusion bodies pellet (P). Subcloning to pET SUMO vector improved the expression level (Fig. 14B) but the SUMO fusion stayed insoluble in the pellet (P). The size of SUMO-fused redesigned Urzyme was at the expected position of approximately 25 kDa. TrpRS Urzyme accounted for 56% of the fusions which suggested that the 11 kDa SUMO carrier did not have a sufficiently large molecular weight to assist proper folding for solubility of the Urzyme. A more powerful carrier, involving a bigger carrier protein was desired after this first vector failed to solubilize the Urzyme. MBP became a reasonable candidate for the next step.

The Majority of MBP Fusion Protein was in the Soluble Fraction

The simplified linearlized pMAL-c2 x map used for subcloning is shown in Figure 16A below. MBP (blue) was approximately three times bigger than the subcloned open reading frame of TrpRS Urzyme (ORF, green). The construct was doubly tagged with N-terminal FLAG and C-terminal 6xHIS. The soluble (S) and insoluble (P) fractions immunoblotted with anti-FLAG antibody showed the majority (80%) of the fusion in the supernatant (Fig. 16B). Amylose resins gave pure, soluble fusion after affinity column (Figure 17). However, the fusion was inactive in the initial tryptophan activation assay (data not shown).

Factor Xa Cleavage Released Tryptophan Activation Activity

To verify the possibility that maltose binding protein, the three times bigger carrier, might interfere with the activity of the smaller TrpRS Urzyme, perhaps by covering its active site, we used Factor Xa according to the manufacturer's manual. Factor Xa, a protease whose recognition site was embedded in between MBP and the

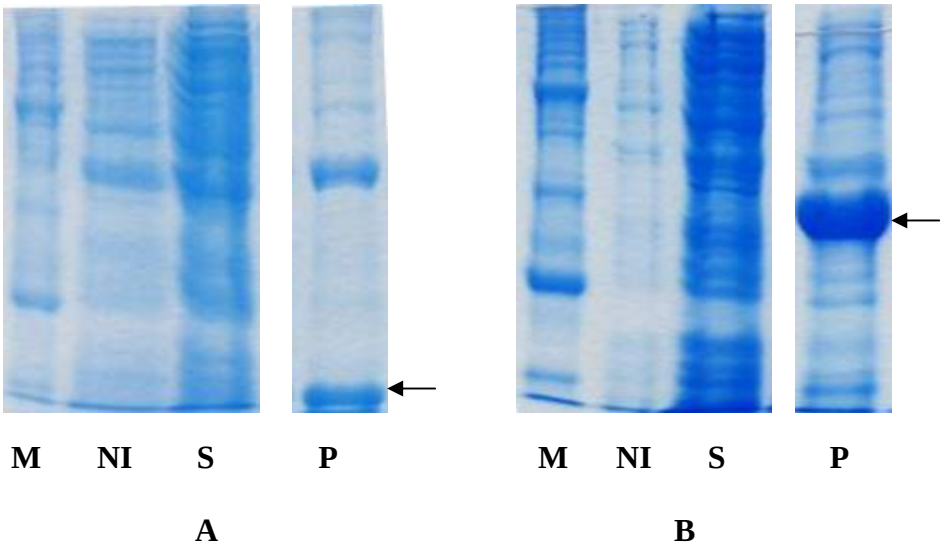


Figure 15: SUMO expression system increases the total Urzyme production but not the solubility (Coomassie Blue stained SDS-PAGE gels). A: redesigned TrpRS Urzyme. B: SUMO fusion. M: molecular weight marker. NI: non-induced control. S: supernatant. P: pellet (Pham, unpublished).

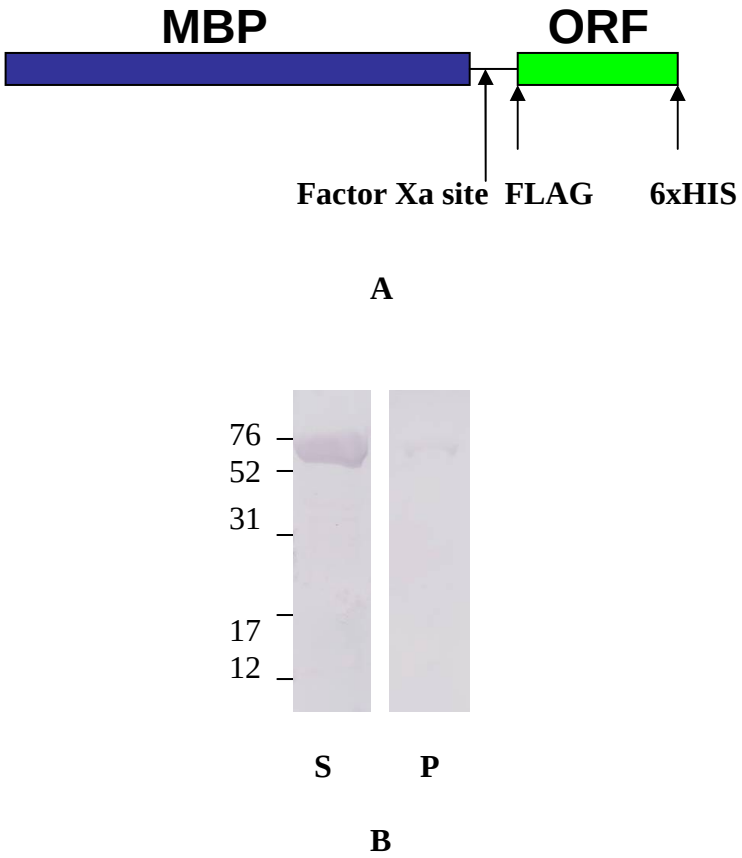


Figure 16: MBP expression system kept the TrpRS Urzyme in the supernatant. A: MBP fusion map with indicated relative positions of the maltose binding protein gene, protease cleavage site, affinity tags and the open reading frame. B: anti-FLAG immunoblotted samples from the soluble (S) and insoluble (P) fractions (Pham, unpublished data).

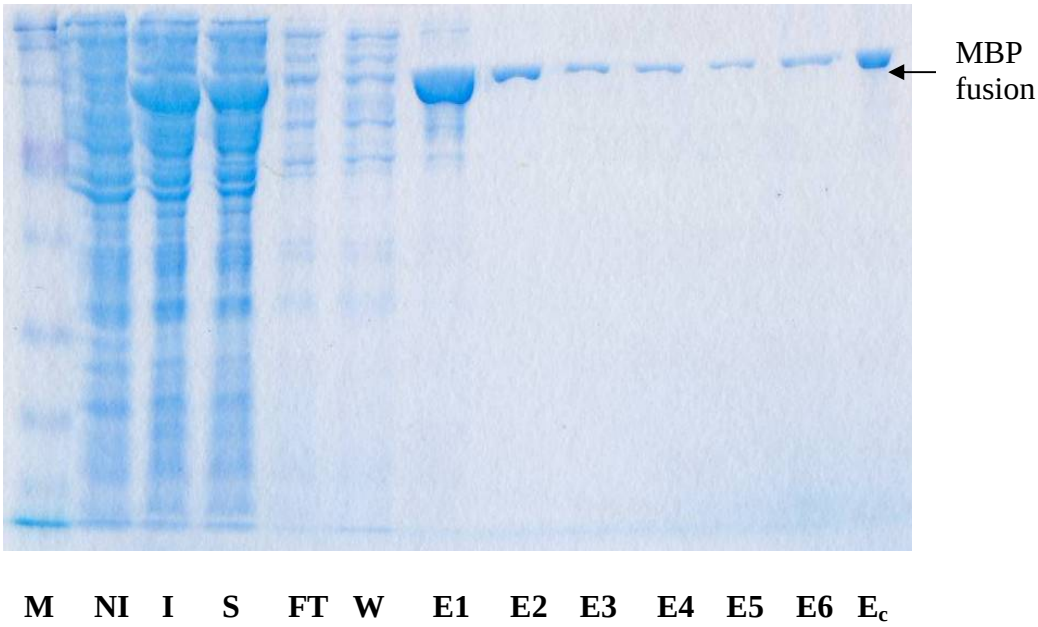


Figure 17: Expression and purification of maltose binding protein-fused redesigned TrpRS Urzyme. M: molecular weight marker. NI: non-induced control. I: induced expression, total fraction. S: induced expression, soluble fraction. FT: flow through fraction from amylose column. W: first washing fraction after FT. E1-6: elution fractions. E_c: concentrated fusion (Pham, unpublished).

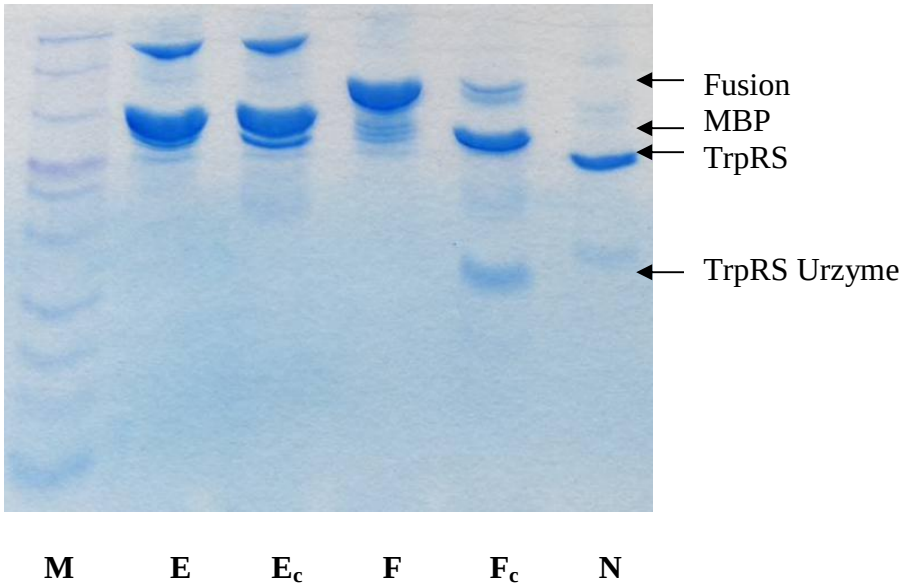


Figure 18: Factor Xa cleavage released the active TrpRS redesigned Urzyme while the same protease did not produce extra digestion products for the empty vector control. M: Molecular weight marker. E: empty pMAL-c2 control. E_c: E cleaved with Factor Xa. F: fusion protein. F_c: F digested with Factor Xa. N: native TrpRS (Pham, unpublished data).

ORF (Fig. 16), was added to digest the fusion protein. As shown in Figure 18 above, the empty vector control (pMAL-c2) (E) was not digested by the protease (E_c) whereas after cleavage, the fusion (F) released the MBP and the Urzyme (F_c). The native full length TrpRS was run in the last lane (N).

The digestion mixture was directly used for pyrophosphate exchange assay. Preliminary results showed that the released Urzyme fragment was active (data not shown).

Discussion

The pET SUMO Expression System did not Solubilize TrpRS Redesigned Urzyme

Marblestone *et al*. (2006) have shown that pET SUMO is powerful in solubilizing proteins which are known to be difficult to express. Besides, the protease used in this system is preferable since it has higher affinity to the substrate and is more active. However, the SUMO-TrpRS Urzyme fusion cannot be successfully expressed in soluble form. The yield is substantially higher but most of the protein aggregates to form inclusion bodies. It is notable that this aggregation also occurs when induction is performed at 18 or 25°C. Under these conditions, *E. coli* does not form inclusion bodies, but the SUMO-TrpRS Urzyme construct is still in the insoluble fraction. It is important to note that the SUMO carrier protein is smaller than the attached TrpRS Urzyme. This phenomenon might indicate that a larger carrier protein is desired to solubilize the highly insoluble redesigned fragment, or, in other words, the SUMO has been overloaded. It is likely that a system working for some proteins will not work for others. Hence, subcloning to a different expression vector is rationalized.

Considerations for Expressing an Enzyme as a Fusion Protein

Since the SUMO fusion seemed to be insufficient to shift the expression into soluble fraction, we tried expression in the pMAL-c2x vector that encodes the maltose binding protein, which is three times bigger than TrpRS Urzyme. This system has successfully solubilized the construct. However, this construct also conceals the Urzyme's activity. A significantly reduced tryptophan activation activity was detected for the fusion form. The observed activity was only 5% compared to that of the cleaved fusion protein preparation.

Yet, when digested with the protease Factor Xa to remove the fused carrier protein, the activity was recovered. This result could be an implication for the

interference of the maltose binding protein with TrpRS Urzyme active site. With a mass three times the size of the Urzyme, it is possible that MBP would make unexpected interactions with residues critical for catalysis. Besides, the carrier might have blocked access of substrates to their binding sites. Hence, the ratio between the sizes of carrier protein and the protein of interest requires considerations, especially when an enzyme is to be attached.

Solved and Remaining Problems

With the implementation of pMAL-c2 expression system, I have successfully solubilized the redesigned TrpRS Urzyme. As previously discussed, the pET42a (+) plasmid carrying TrpRS Urzymes tends to be unstable and lost after some bacterial division cycles, implying possible toxicity. Thus, this soluble fusion form has an advantage over expressing the Urzyme itself regarding the maintenance of the plasmid. Furthermore, the problem of aggregation can be also reduced as long as the Urzyme is fused with the maltose binding protein carrier which probably will increase the storage time of purified material in 50% glycerol at -20°C. The fact that the activity is released after Factor Xa cleavage may be another advantage considering MBP a protecting cage that helps preserve the Urzyme. This protecting role was verified by a digestion experiment done with time, temperature and glycerol concentration as variables in which aggregation was observed after longer incubation time, at higher temperature or without glycerol (data not showed).

It is important to note that the cleavage reaction mixture is used directly for pyrophosphate exchange assay. This condition might interfere with the activity of the Urzyme, since there are remaining reactants and unwanted products (i.e. fusion protein, protease, MBP). Further manipulation is necessary to purify the TrpRS Urzyme after cleavage. However, as noted next, several failures to produce activity

after cleavage of the fusion protein confirm that none of these contaminants has competing activity.

A different problem recently identified is that a second lot of Factor Xa (NEB) seems to produce intriguing cleavage pattern with extra bands (data not shown), and failure to reproduce tryptophan activation activity. This could be a sign for non-specific internal cleavage since there is no Ile-Glu-Gly-Arg sequence (the protease's canonical recognition site) in the TrpRS Urzyme. Investigation is to be done to explain, and hopefully solve, this remaining problem.

Conclusion

Of the two expression systems used to solubilize the redesigned TrpRS Urzyme, pET SUMO produced substantial amount total protein compared to the original pET42a; however, it was not successful in keeping the construct in the supernatant fraction. A higher molecular weight carrier protein, MBP, was chosen to provide stronger solubilizing force. MBP-fused Urzyme was soluble but did not have expected tryptophan activation activity. Factor Xa, the protease used to separate the carrier protein MBP and the protein of interest, released the Urzyme fragment which showed activity in the pyrophosphate exchange assay. Further characterization of catalysis by this Urzyme will help unravel the role of the missing domains, and is discussed in Chapter 4.

Acknowledgement

This work is supported by NIH grant GM078227-01. pMAL-c2x vector was a kind gift from Dr. Sancar. I would like to thank N. Ozturk for assisting with the set up of MBP system to efficiently express and purify the fusion protein. I also would like to give special thank to Dr. Fried for providing advice and other reagents to accomplish the goal of solubilizing the construct.

CHAPTER 4

TRYPTOPHANYL-TRNA SYNTHETASE URZYME SHOWS COMPARABLE AFFINITY FOR ATP BUT A 10^3-FOLD HIGHER K_M FOR TRYPTOPHAN IN AMINO ACID ACTIVATION, INDICATING A POTENTIAL ROLE FOR THE CONNECTING PEPTIDE (CP1) IN AMINO ACID BINDING POCKET STABILIZATION AND AMINO ACID SPECIFICITY

Abstract

A stable and soluble tryptophanyl-tRNA Urzyme has been expressed in the fusion form with maltose binding protein in the pMAL-c2x vector system. This Urzyme is the result of a computational design phase in which the connecting peptide 1 and the C-terminal domain were removed. 17 mutations were introduced in the construction to stabilize and solubilize the Urzyme. The preliminary results show significant tryptophan activation activity for this 130-residue long peptide. In this chapter, Michaelis-Menten kinetics parameters for the TrpRS Urzyme obtained from $^{32}PP_i$ phosphate exchange assay will be discussed. Compared to the full length enzyme, TrpRS Urzyme binds ATP with similar affinity; however, tryptophan binding is approximately 10^3 times weaker. This result argues that either the removed CP1 and CTD domains, or both, contribute in important ways to amino acid specificity. Furthermore, since the TrpRS CP1 insertion occurs at the same position as the editing domain of five other aaRSs, this finding is important evidence supporting the conclusion that the CP1 fragment has always served to enhance the amino acid specificity of Class I aaRS. CP1 therefore appears to be the potential birthplace for editing activity, which would also support the hypothesis of later internal insertion to a primitive polypeptide (Rodin and Ohno, 1995; Pham et al., 2007).

Introduction

AaRSs very likely evolved from a simpler, more primitive polypeptide chain into complex multi-domain, multi-function contemporary enzymes with enhanced catalytic activity and fidelity. They have been shown to participate in various biological processes. Structural analyses indicate that both class I and class II aaRSs contain small (~120 amino acids) catalytic cores (Rodin and Ohno, 1995). Complementing these cores, there are internal and external insertions, some well characterized structurally and functionally, the others still under investigation. For instance, the N-terminal extension of mammalian TrpRS is subjected to proteolysis, producing natural mini TrpRS variants shown to have antiangiogenic/cytokine activity. The carboxyl-terminal domain of aaRS interacts with the anticodon, which is important for RNA recognition and binding affinity. More interestingly, a domain called connecting peptide 1, which varies in size and in five aaRSs, is shown to be the originating location of editing activity that amplifies the fidelity of aminoacylation in those cases involving similar naturally occurring amino acids.

There are two proposed mechanisms for aaRS editing: pre-transfer in which the non-cognate adenylate is hydrolyzed and post transfer that works on clearing mis-acylated tRNA through hydrolysis (Schimmel and Schmitdt, 1995). Each mechanism has distinct features and varies from one aaRS to another. Since pre-transfer editing occurs in the first half of the two-step aminoacylation reaction, tRNA might not be required for the removal of mis-activated amino acids. However, for some aaRSs, e.g. LeuRS, it has been shown that the canonical core and CP1 both contribute to the proofreading activity (Boniecki *et al*., 2008; Zhou *et al*., 2008). Homologs of CP1 have been found to express editing activity in a number of organisms. The *in trans* co-existence of ProX and YbaK with bacterial ProRS and CysRS, respectively, enhances

the fidelity of aminoacylation (Ahel *et al*., 2003; An and Musier-Forsyth, 2005). At first glance, these isolated domains seem to be redundant. However, this editing activity provided from a different protein might be a primitive mechanism before domain fusion to the catalytic core brought the non-cognate substrates into closer proximity to the editing site.

Evidence for evolution at the molecular level has been well documented for a number of enzymes. A frequently observed and proved mechanism in this process is gene duplication followed by domain fusion or insertion (Traut, 2008). TrpRS CP1 is only 74-residue long and the native full length enzyme does not have editing activity. However, it is an ideal system to study since it is the smallest aaRS and the gap between the two free ends resulted from CP1 removal is only 5 Ångstrom long, which can be joined by a short linker combined with relaxed adjacent backbone. A molecule of approximately 120-130 residues in length like TrpRS Urzyme might be a potential candidate for the ancestral class I aaRS. Under this hypothesis, other additional domains were later insertions and fusions to the primordial genes. The activity of TrpRS Urzyme previously described in Chapter 2 (published as Pham *et al*., 2007) supports this paradigm. However, further characterization is necessary to determine the detailed mechanism and explore the role of the removed domains CP1 and CTD.

The idea of ancestral gene resurrection has been implemented by several groups to study evolution at the molecular level. One of the most striking results comes from the work on the steroid hormone receptors (Thornton and Bridgham, 2007). In Chapter 16 of the book Ancestral Sequence Reconstruction, the authors have described their mutational analysis with the contemporary protein. By making multiple point mutations, they were able to modulate the specificity of the receptor for different ligands. Combined with the divergence of ligand specificity along the

phylogenetic tree, the group has built a detailed pathway for the receptor evolution, further indicating that some of the evolutionary events were irreversible (Bridgham, Ortlund and Thornton, 2009). This TrpRS Urzyme construction, pursuing the same target of ancestral gene resurrection to study the recapitulation of putative evolutionary events, is extremely challenging because it entails removal of whole structural domains to create a model for the ancestral enzyme, whose activity can be measured. On this system, changes can be made that might have occurred along the evolution leading to the contemporary enzyme. It is especially important that the effects of those changes can also be measured. In contrast to ancestral gene resurrection, however, amino acid changes in the conserved domains cannot be identified from a phylogenetic sequence analysis; we use biophysical chemistry from protein design to approximate putative sequence changes (Pham *et al.*, 2007).

A common steady state measurement used to study enzyme kinetics is the Michaelis-Menten experiment. Proposed in 1913 by the two scientists that this method is named after, Michaelis-Menten equation has been a standard representation of the relationship between an enzyme and its substrate (Michaelis and Menten, 1913).

$$v = v_{max}[S]/(K_m + [S]) \quad (I)$$

$$\text{such that } v_{max} = k_{cat}[E_o] \quad (II)$$

Fitting the concentration-dependent activity data to equation (I) is used to determine two parameters: the maximum velocity v_{max} and the binding constant K_m which are critical to understand the kinetics of any enzyme. From the main equation, the apparent second-order rate constant (the ratio of k_{cat} over K_m) of the enzyme can also be determined. Use of this parameter is a widespread indicator for enzymatic proficiency. Another assay to be considered is active site titration (Fersht, 1999) to

determine the active proportion of the total enzyme used; active-site titration is essential for accurate determinations of k_{cat} because of relation (II). We have not yet performed active-site titration on the TrpRS Urzyme. This chapter will be devoted to preliminarily study the Michaelis-Menten kinetics of tryptophan activation, which is the first half of the reaction catalyzed by the full length TrpRS enzyme.

Materials and Methods

Expression and purification of MBP fusion

MBP-TrpRS Urzyme fusion was expressed directly after transforming with plasmid DNA. After transformation, 100 µl competent cells were incubated in 1ml LB media at 37°C for 1 hr with shaking and then used to inoculate 100 ml growth media (LB plus 2% glucose, 100 mg/ml for ampicillin). The inoculum was shaken at 250 rpm overnight at 37°C and then transferred to 1 l fresh LB with 100 mg/ml ampicillin. At OD_{600} = 0.4–0.6, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 0.3 mM final concentration to induce gene expression. Harvest began 3 hr after induction.

Cells were collected by centrifugation at 4,500 rpm for 20 min at 4°C, and then resuspended in lysis buffer (50 mM Tris, 10% sucrose, pH 7.5). After overnight storage at -80°C, the cells were thawed on ice and lysed by sonication (3x15 sec). The lysate was centrifuged to separate the supernatant fraction from the inclusion bodies at 14,000 rpm for 30 min at 4°C. The supernatant from this step was diluted 1:5 with the amylose resin High Flow (NEB) column buffer (20 mM Tris-HCl, 0.2 M NaCl ,1 mM EDTA,10 mM beta-mercaptoethanol, pH 7.4) before loading onto a calibrated Biorad disposable column contained 2 ml of amylose-conjugated resin suspension for each 10 ml of the diluted supernatant. The column was equilibrated with 3 column volumes (c.v.) of column buffer and the resin was washed with 5 c.v. The flow-through fraction was collected then the resin was washed with 8 c.v. of column buffer to minimize non-specific binding. The bound fusion protein was eluted in 1.5 ml fractions with elution buffer (column buffer plus 10 mM maltose). The first five fractions were combined and dialyzed overnight against 3x10 volumes of 50% glycerol dialyzing buffer (20 mM Tris-HCl, 50 mM KCl, 0.1 mM PMSF, 10 mM

beta-mercaptoethanol, 50% glycerol, pH 7.5) at 4°C. Samples were stored at -20°C after dialysis. An example of the purity of this preparation is shown in Chapter 3, Figure 17.

Factor Xa Cleavage

The MBP fusion was digested with Factor Xa (NEB) overnight at 4°C with rotating. The ratio of substrate to enzyme is 1:50 (w/w). Concentrated fusion protein was diluted to a final concentration of 10% glycerol from the stock solution stored in 50% glycerol (1:5 dilution). The digestion mixture was used directly for the Michaelis-Menten tryptophan activation assays.

ATP- and tryptophan- dependent ^{32}PPi Exchange Assays for Tryptophan Activation

Samples were assayed using the traditional ^{32}P-pyrophosphate exchange as described (Joseph and Muench, 1971) except that a 3-fold reduction of background ^{32}P counts was achieved by collecting the charcoal containing labeled ATP on disposable spin columns, washing, and eluting the bound ATP with 50 µl of pyridine at 37°C. ATP and tryptophan were first depleted from the assay buffer in order to determine Michaelis-Menten constants as ATP- and tryptophan- dependent parameters, respectively. Stocks of ATP and tryptophan were added to the depleted buffer at indicated ranges of concentrations. Activity measurements collected were plotted against substrate concentrations.

Results

TrpRS Urzyme Tryptophan Activation Activity is Dependent on ATP

The k_{cat}/K_m previously obtained from refolded, redesigned TrpRS Urzyme expressed in pET42a plasmid was 10^5 times weaker than that of the full length enzyme. Digestion with Factor Xa recovered a cryptic tryptophan activation activity in the fusion protein. As described in the Methods section, to characterize ATP-dependent Michaelis-Menten kinetics, ATP concentrations ranging from 0 up to 2 mM were prepared. In order to compare such widely different enzymatic activities, the ratio v/v_{max} values were plotted against ATP concentration (Fig. 19). There was no activity when ATP was absent (concentration = 0 mM). As the ATP concentration increased, as shown in Figure 18 below, the kinetics of tryptophan activation of the native full length enzyme (dark blue diamond) was similar to that of the redesigned Urzyme (magenta square). They both showed typical hyperbolic saturation curves with v_{max} being approached at ATP concentration of approximately 2 mM.

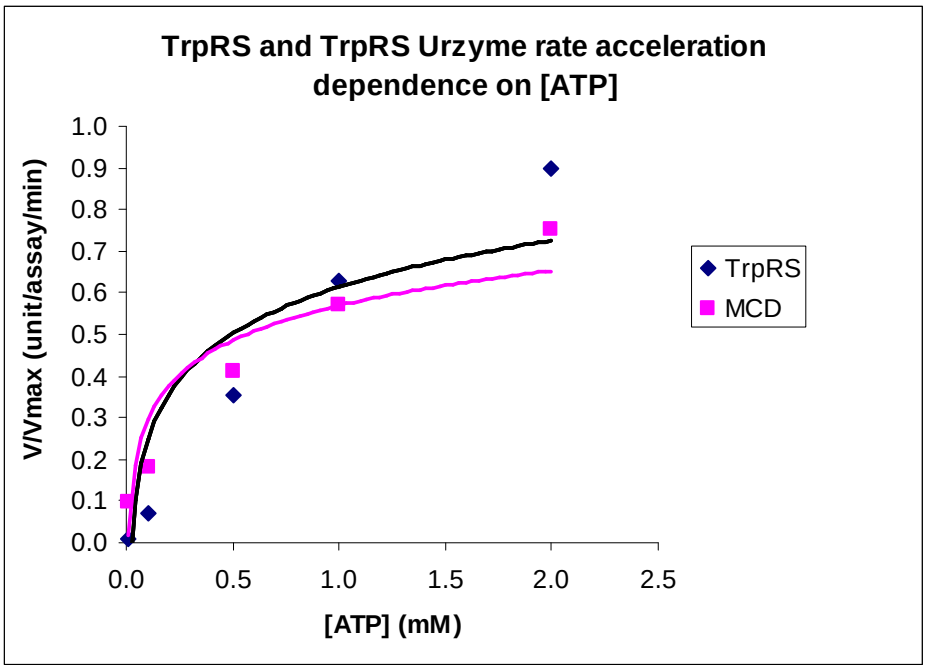
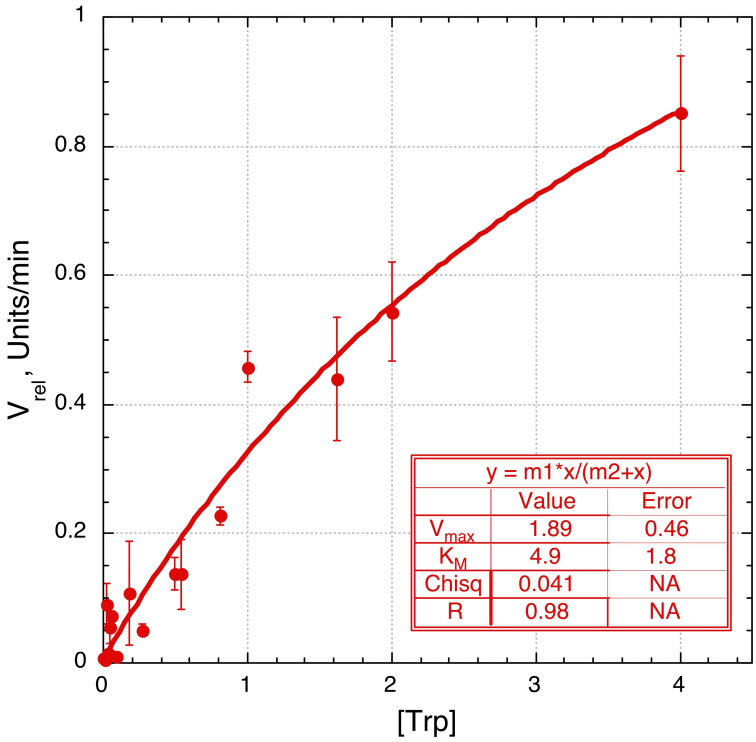


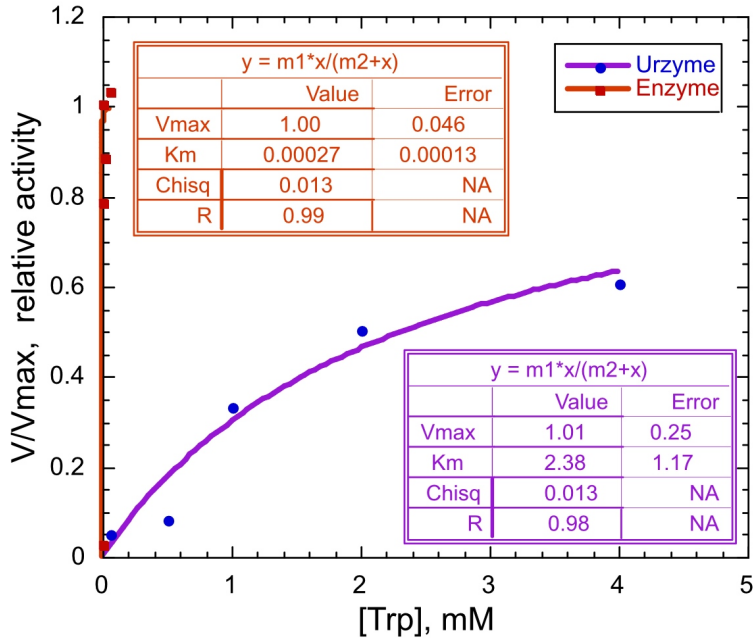
Figure 19: ATP dependence of TrpRS and TrpRS Urzyme (MCD – minimal catalytic domain) in rate acceleration of tryptophan activation (Pham, unpublished).

Tryptophan Dependent Activity Assay

A similar experiment was done to determine the tryptophan dependence of TrpRS Urzyme in tryptophan activation. As expected, when tryptophan was absent, there also was no activity detected. However, the Urzyme became more active as more tryptophan was added to the assay buffer. The saturation curve was not yet completely observed at 4 mM of the cognate amino acid substrate (Fig. 20A). On the other hand, the full length TrpRS reached maximum velocity rather early within the experimental tryptophan concentration range (Fig. 20B, orange line). When v/v_{max} ratios were plotted as before against different tryptophan concentrations used to titrate the Urzyme activity, the curve was essentially vertical. The behavior observed for the TrpRS Urzyme indicates a quantitatively different, and impaired tryptophan binding affinity for the TrpRS Urzyme which we verified by fitting the observed data to obtain Michaelis-Menten constants shown in Figure 20 and Table 7.



A



B

Figure 20: **Full length enzyme and the Urzyme activate tryptophan differently.** **A**: Normalized tryptophan activation by the redesigned TrpRS Urzyme. **B:** Normalized tryptophan-dependent rate acceleration by full length enzyme (red) and Urzyme (blue) (Pham, unpublished).

TrpRS Urzyme has Comparable ATP Binding Affinity but Approximately 10^3 Times Weaker Trp Binding Affinity Compared to the Full Length Enzyme

The Michaelis-Menten constants obtained are summarized in Table 7 below. Compared to the native TrpRS, the Urzyme bound ATP equally tightly with a K_m of 0.68 mM. Rate acceleration was substantially lower which was consistent with previously determined activity for the renatured Urzyme. Interestingly, tryptophan binding affinity of the construct was greatly reduced. The K_m value for tryptophan was approximately 1000 times higher (Table 7, bold), which indicated that the Urzyme might not be specific for its cognate substrate.

	Full length TrpRS	TrpRS Urzyme
K_m for ATP (mM)	0.68	0.70
V_{max} (ATP)	0.054	0.0001
K_m for Trp (mM)	**0.00027**	**2.38**
V_{max} (Trp)	0.034	0.0001

Table 7: Michaelis-Menten parameters determined for full length TrpRS and the redesigned TrpRS Urzyme (Pham, unpublished).

Discussion

TrpRS Urzyme Binds ATP with Affinity Comparable to the Full Length Enzyme

The TrpRS Urzyme is designed by fusing the two discrete halves of the ATP-binding Rossmann fold. Its ATP binding site composed of the TIGN and KMSKS signature loops is retained in the *in silico* construction phase. Hence, binding affinity to ATP is expected. However, the fact that this severely truncated peptide binds ATP as tight as the full length enzyme is striking. It is interesting that a TrpRS 46-mer peptide constructed from the first 46 amino acids of TrpRS containing only TIGN loop binds ATP even more tightly (L. Li, V. Weinreb, unpublished data). Since binding to ATP is the first step of the aminoacylation reaction, this feature is important to be further characterized. More interestingly, there is evidence showing that high concentration of ATP eventually destabilizes the full length enzyme (Carter, data not published). This behavior has not been tested with the Urzyme. Furthermore, the widely accepted K_m value for ATP of the native TrpRS is 0.4 mM which is lower than the apparent value of approximately 0.7 mM discussed earlier in this chapter (Table 7). It is likely that the enzyme used in the assay was more active. Nevertheless, future assays are necessary to verify the difference.

Binding Affinity toTtryptophan is Approximately 1000 Times Weaker

The dramatic difference between the TrpRS Urzyme and full-length TrpRS is the weaker binding of tryptophan by the Urzyme. As shown in Table 10, whereas the ATP binding affinity remains essentially unchanged, the tryptophan affinity is approximately 10^3-fold different. The interpretation of this is that the redesigned Urzyme is unable to bind tryptophan efficiently. This result is unexpected considering the fact that the tryptophan binding pocket is preserved like the ATP binding site. However, a likely possibility is that some long-range interactions might have been

eliminated by removing the connecting peptide 1 and the carboxyl-terminal anticodon binding domain. This significantly reduced affinity could implicate *a potential role for the connecting peptide 1 in amino acid recognition* (Fig. 21 below).

Evolution from a Primitive Polypeptide by Acquiring Functional Domain through Gene Fusion and Insertion Supports Rodin-Ohno Hypothesis

Fig. 21 represents the resurrected connecting peptide 1 in a relative position to the redesigned TrpRS Urzyme. As previously mentioned in the introduction of this chapter, the TrpRS CP1 insertion is small and the native TrpRS does not have editing activity; nevertheless, at the same location has grown a bigger, structured domain capable of clearing mis-aminoacylated products in four other aaRSs (MetRS, LeuRS, IleRS, ValRS). More interestingly, the gap between TrpRS residues T46 and G121 after its CP1 removal is only 5 Ångstrom long, indicating a possibility for ancestral direct linkage. It can be seen from the above figure that CP1 surrounds the tryptophan binding pocket. The insertion may work as an exo-skeleton to form and stabilize the proper configuration necessary for tryptophan recognition and binding.

These results suggest a rather unexpected view of the ancestral TrpRS Urzyme, in which ATP bound tightly but tryptophan had only modest affinity. More importantly, the resutls bring the implications of low amino acid specificity into the discussion. It is interesting to note that contemporary TrpRS is very specific for its cognate amino acid substrate, tryptophan. Praetorius-Ibba *et al.* (2000) made several mutations within the tryptophan binding pocket in an effort to switch its specificity to tyrosine. However, their conclusion confirms the failure to modulate amino acid specificity for this class Ic aaRS, indicating a possibility for the contribution of longer range interactions in defining the tryptophan binding pocket.

Recently, similar attemps to change GlnRS binding specificity from glutamine to glutamate have confirmed the importance of longer range interactions in configuring the amino acid binding pocket (Bullock *et al*., 2008). The authors show that introduction of Glu primary binding sites only into GlnRS is insufficient for the specificity switch. However, additions of two more engineered loops and one deletion drastically increase the generation of misacylated Glu-tRNA^{Gln} by GlnRS with 16,000 fold difference. Interestingly, further investigation of the glutamine binding pocket reveals unusual interactions of the amino acid with ATP, the second substrate in the first step of the aminoacylation reaction (Corigliano and Perona, 2009). These studies on the class Ib GlnRS strongly support the existence of distal interactions outside of the tryptophan binding pocket in TrpRS, possibly originating from the CP1 domain.

Early proteins were probably non-specific molten globules and hence we might expect that a non-specific type I aaRS would have had a selective advantage because it could generate a larger number of potentially functional variants. Only later, as the genetic code became firmly established, and globular proteins became more specific, would it have been necessary to evolve higher amino acid specificity. Multiple events of duplication, fusion, and insertion probably brought the additional domains assisting the primitive enzyme in substrate affinity and fidelity enhancements. The result of this postulated divergent evolution is the twenty contemporary aaRSs which are remarkably different, regarding both structure and mechanism. However, it was certainly essential for any imaginable protein synthesis machinery that ATP be bound tightly and productively. The native-like ATP binding affinity observed for the TrpRS Urzyme in tryptophan activation might be supportive of its becoming a potential candidate for the ancestral Urzyme.

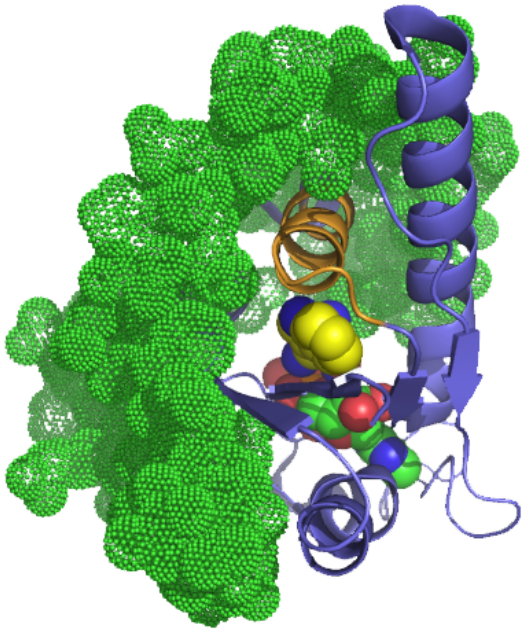


Figure 21: ***In silico*** **resurrection of the connecting peptide 1 (green spheres) wrapping around the core TrpRS Urzyme that binds tryptophan in the amino acid binding pocket** (Carter, unpublished).

Conclusion

The solubilized redesigned TrpRS Urzyme tryptophan activation activity is further characterized through determination of Michaelis-Menten constants. It is striking that this 130 amino acid long peptide possesses a native-like ATP binding affinity but a 10^3-fold weaker affinity to the cognate substrate, tryptophan. This indicates that the removed domains might be crucial for amino acid specificity. More importantly, considering the TrpRS Urzyme a potential candidate for the ancestral enzyme and CP1 and CTD later insertions in evolution, the addition of these two domains would enhance the observed weak affinity for tryptophan. It is possible that the primitive Urzyme was able to bind ATP tightly but unable to distinguish its cognate amino acid efficiently. Activation of non-cognate amino acids, especially the structurally closely related tyrosine, either in the absence or presence of CP1 and CTD, is to be further investigated.

Acknowledgement

I would like to thank all members in Carter group, L. Li, V. Weinreb, and O. Erdogan, for their supports and encouragements. The TrpRS 46-mer is related to O. Erdogan's work. L. Li has made and characterized the TrpRS 46-mer. V. Weinreb has helped with the pyrophosphate exchange assay. I also would like to thank Dr. Fried for invaluable suggestions while working on this project.

This work was partially supported by the NIH supplemental grant and would be prepared to publish.

CHAPTER 5

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Aminoacyl-tRNA synthetases participate in many crucial biological processes employing different activities beside their canonical aminoacylation role in protein synthesis. These enzymes can be potential targets to treat patients suffering from a variety of diseases including cancer, Alzheimer, autoimmunity, etc. It is important to study the intramolecular interactions within the enzyme to understand the catalytic mechanism. Furthermore, investigating the question of how two contemporary class I and class II aaRS might have emerged simultaneously and then evolved separately would shed light on the involvement of these enzymes in current pathways and perhaps allow the prediction of new functions.

Chapter 2 describes a TrpRS Urzyme constructed by fusing two discrete halves of the ATP binding Rossmann fold that shows significant amino acid activation activity. This is a uniquely challenging work since instead of making point mutations in the intact enzyme, an Urzyme is carved out with extensive deletion and truncation. The 130-residue long peptide, lacking the connecting peptide 1 and the carboxyl-terminal anticodon binding domain that make up 60% the mass of the active dimer, activates tryptophan with 65% of the full strength rate acceleration potential (see Chapter 2 conclusion). It is important to know that this is not contaminating activity from the host cell TrpRS since the empty vector control has been substracted as background noise. Besides, TrpRS Urzyme has been purified through renaturation from the inclusion bodies, separately from the soluble fraction that could contain native form of the enzyme. Further experiments are to be done to characterize the Urzyme to understand the modular complementation of the removed domains. To that end, I have prepared the B, D, ABC and ABD fragments. I will use these to test the influence of CP1 and CTD on specificity and activity.

More importantly, this Urzyme is an excellent potential candidate for the ancestral class I aminoacyl-tRNA synthetase postulated in the Rodin-Ohno hypothesis. The procedure I have developed in constructing the TrpRS Urzyme can be used as the model to construct other class I Urzymes, as well as a class II Urzymes. Such experiments are underway in the lab, and can provide important data to verify the validity of the hypothesized sense/antisense relationship.

Chapter 3 describes the two expression systems used to solubilize the redesigned TrpRS Urzyme, in which the pET SUMO produced a substantial amount of total protein compared to the original pET42a; however, it was not suitable for obtaining soluble protein. A higher molecular weight carrier protein, MBP, was chosen to provide stronger solubilizing force. MBP-fused Urzyme was soluble but did not have expected tryptophan activation activity. Factor Xa, the protease used to separate the carrier protein MBP and the protein of interest, released the Urzyme fragment which showed activity in the pyrophosphate exchange assay. Further characterization of catalysis by this Urzyme will help unravel the role of the missing domains, and is discussed in the next chapter.

Michaelis-Menten constants were determined through assaying the Urzyme with increasing concentrations of the two substrates, ATP and tryptophan. The exchange assay assures that the K_m values for both substrates actually represent thermodynamic affinities. The results with both substrates are somewhat surprising. Binding affinity to ATP is relatively unchanged; however, K_m for tryptophan is 10^3 fold weaker, indicating that the amino acid binding pocket is impaired in the Urzyme. This is probably the result of removing connecting peptide 1, which might have distal interactions with the tryptophan binding cavity and perhaps also the GxDQ loop. A potential role for CP1 and CTD will be further investigated by *in cis* and *in trans* co-

expression of the two domains ideally with the wild type Urzyme, which would help understand the detailed steps in the enzyme evolution. More importantly, TrpRS Urzyme can serve as a model system to construct a class II aaRS Urzyme to verify the hypothesis proposed by Rodin and Ohno about two coding strands for both classes from one ancestral gene.

The discussed results in the above projects are solely based on the activity assay for the amino acid activation. As previously described, after this initial step, the intact contemporary enzyme will catalyze the acylation of its cognate tRNA. Wolfson and Uhlenbeck have established an assay to characterize the second half of the aminoacylation reaction, which is necessary to understand the complete catalytic mechanism by TrpRS Urzyme. It is expected that the signal, if detectable, would be weak since most of the known structural RNA-interacting elements have been removed. Further investigation will be done on which form the construct might accommodate. This question is interesting since the dimer interface required for dimerization has been deleted together with the connecting peptide 1.

I have constructed the connecting peptide 1 and the C-terminal anticodon binding domain, as well as two genes encoding these domains *in cis* relative to the core Urzyme. These four constructs are to be characterized to unravel the intramolecular interactions within the full length enzyme. More importantly, this experiment can verify the predicted modular complementation of CP1 and CTD regarding substrate recognition and binding enhancement. Besides the *in cis* co-expression, it seems that we should be able to redesign the free CP1 and CTD to test their activity *in trans* to study the nature of the complementation to see whether it is covalent or non-covalent. Similar experiments have been done with the GluRS (Lapointe *et al*., 2009), in which the authors have put the truncated CTD back *in trans* and shown that the activity is

enhanced. It is likely that TrpRS Urzyme might have acquired CP1 and CTD as additional domain insertion later in evolution. This would further support Rodin-Ohno hypothesis about the existence of a small primitive ancestral class I enzyme.

This class I Urzyme, according to the sense/antisense hypothesis, would have a complementary class II Urzyme. There are two different experiments that can be done to test this complementarity. First, based on the gene sequence of TrpRS Urzyme, an antisense gene will be constructed. The corresponding peptide will be expressed if possible then assayed for amino acid activation activity. The second experiment, which is more feasible, is to construct a class II Urzyme from the highly conserved motifs of this class then characterize the fragment for aminoacylation activity. The Urzyme can be used as a model system to build an ancestral catalytically active core, on which presumed later insertions can be added and evaluated for their contributions. This work on defining possible intramolecular interactions among structural domains would complement the point mutational analyses in understanding evolution at the molecular level.

REFERENCES

Ahel, I., Korensic, D., Ibba, M., and Soll, D. (2003). Trans-editing of mischarged tRNAs. *Proc. Natl. Acad. Sci.* 100(26):15422-27.

An, S. and Musier-Forsyth, K. (2005). Cys-tRNA^{Pro} editing by Haemophilus Influenzae YbaK via a novel synthetase·YbaK·tRNA ternary complex. *J. Biol. Chem.* 280:34465-72.

Augustine, J. and Francklyn, C. (1997). Design of an active fragment of a class II aminoacyl-tRNA synthetase and its significance for synthetase evolution. *Biochemistry* 36(12):3473-3482.

Boniecki, M.T., Vu, M.T., Betha, A.K., and Martinis, S.A. (2008). CP1-dependent partitioning of pretransfer and posttransfer editing leucyl-tRNA synthetase. *Proc. Natl. Acad. Sci*. 105(49):19223-28.

Bridgham, J.T. and Thornton, J.W. (2007). Ancestral sequence reconstruction (Book). Chapter 16:183-200.

Bridgham, J.T., Ortlund, E.A., and Thornton, J.W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515-519.

Bullock, T.L., Rodríguez-Hernández, A., Corigliano, E.M., Perona, J.J. (2008). A rationally engineered misacylating aminoacyl-tRNA synthetase. *Proc. Natl. Acad. Sci.* 105(21):7428-33.

Burbaum, J.J. and Schimmel, P. (1991). Structural relationships and the classification of aminoacyl-tRNA synthetases. *J. Biol. Chem* 266(26): 16965-16968.

Burbaum, J.J., Starzyk, R.M., Schimmel, P. (1990). Understanding structural relationships in proteins of unsolved three-dimensional structure. *Proteins.* 7(2):99-111.

Campbell-Burk, S.L., Carpenter, J.W. (1995). Refolding and purificationi of Ras proteins. *Methods Enzymol.* 255:3-13.

Carter, C. W., Jr. (1993). Cognition, Mechanism, and Evolutionary relationships in aminoacyl-tRNA synthetases. *Ann. Rev. Biochem.* 62: 715-748.

Carter, C. W., Jr. and Duax, W. L. (2002). Did tRNA synthetase classes arise on opposite strands of the same gene? *Molecular Cell* 10: 705-708.

Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J. C., and Moras D. (1994). The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. *EMBO J.* 13(2): 327–337.

CCP4 (1991). The SRC(UK) Collaborative computing project No 4: A suite of programs for protein crystallography. Daresbury, U.K, Daresbury Laboratory.

Chester, A., Weinreb, V., Carter, C.W., Jr., and Navaratnam, N. (2004). Optimization of apolipoprotein B mRNA editing by $APOBEC_1$ apoenzyme and the role of its auxilary factor, ACF. *RNA* 10:1399-1411.

Corigliano, E. and Perona, J.J. (2009). Architectural underspinnings of the genetic code for glutamine. *Biochemistry* 48:676-87.

Current Protocols in Molecular Biology (Book). (1991). Wiley Intersciences.

Cusack S. (1993). Sequence, structure and evolutionary relationships between class 2 aminoacyl-tRNA synthetases: an update. *Biochimie* 75:1077-1081.

Dantas, G., Kuhlman, B., Callender, D., Wong, M. and Baker, D. (2003). A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* 332(2): 449-460.

Dantas, G., Watters, A.L., Lunde, B.M., Eletr, Z.M., Isern, N.G., Roseman, T., Lipfert, J., Doniach, S., Tompa, M., Kuhlman, B., Stoddard, B.L., Varani, G., Baker, D. (2006). Mis-translation of a computationally designed protein yields an exceptionally stable homodimer: Implications for protein engineering and evolution. *J. Mol. Biol.* 362(5): 1004-1024.

Dean, A.M. and Thornton, J.W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. Nature Reviews Genetics 8:675-688.

Dubois, D.Y., Blais, S.P., Huot, J.L. and Lapointe, J. (2009). A C-truncated glutamyl-tRNA synthetase specific for $tRNA^{Gln}$ is stimulated by its free complementary distal domain: mechanistic and evolutionary implications. *Biochemistry* 48(25):6012-21.

Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature (London)* 347: 203–206.

Fersht, A. (1999). Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding (Book). Chapter 4(E):155-158.

Francklyn, C.S. (2008). DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression. *Biochemistry* 47(45): 11695-703.

Harper, S. Speicher, D.W. (2008). Expression and purification of GST fusion proteins. *Curr. Protoc. Protein Sci.* Chapter 6: Unit 6.6.

Hartman, M. C., Josephson, K., and Szostak, J. W. (2006). Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci* 103(12):4356-4361.

Hausmann, C.D. and Ibba, M. (2008). Structural and functional mapping of the archaeal multi-aminoacyl-tRNA synthetase complex. *FEBS Letters* 582:2178-2182.

Hausmann, C.D., Praetorius-Ibba, M. and Ibba, M. (2007). An aminoacyl-tRNA synthetase:elongation factor complex for substrate channeling in archaeal translation. *Nucleic Acids Research* 35:6094-6102.

Ibba, M., Francklyn, C. and Cusack, S. (2005). *Aminoacyl-tRNA Synthetases*. MBIU, Landesbioscience, Georgetown, TX.

Jakubowski, H. and Goldman, E. (1992). Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Mol. Biol. Rev.* 56(3):412-429.

Jia, J., Arif, A., Ray, P., Fox, P. (2008). WHEP domains direct noncanonical function of glutamyl-prolyl-tRNA synthetase in translational control of gene expression. *Mol. Cell* 29(6):679-690.

Jia, J., Xu, F., Chen, X., Chen, L., Jin, Y., and Wang, D. T. (2002) Two essential regions for tRNA recognition in *Bacillus subtilis* tryptophanyl-tRNA synthetase. *Biochem. J.* 365: 749-756.

Joseph, D.R. and Muench, K.H. (1971). Tryptophanyl transfer ribonucleic acid synthetase of Escherichia coli. *J. Biol. Chem.* 246(24):7602-7609.

Kirby, A. J. and Younas, M. (1970). The reactivity of phosphate esters. Reactions of diesters with nucleophiles. *Journal of the Chemical Society* B 418: 1165-1172.

Kovaleski, B. J., Kennedy, R., Hong, M. K., Datta, S.A., Kleiman, L., Rein, A., Musier-Forsyth, K. (2006). In vitro characterization of the interaction between HIV-1 Gag and human lysyl-tRNA synthetase.*J. Biol. Chem.* 281(28):19449-56.

Kuhlman, B. and Baker, D. (2004). Exploring folding free energy landscapes using computational protein design. *Curr. Opin. Struc. Biol.* 14(1): 89-95.

Kumar, R.K., Yarus, M. (2001). RNA-catalyzed amino acid activation. *Biochemistry* 40(24):6998-7004.

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility *J. Mol. Biol.* 55: 379 – 400.

Lee, J.W., Beebe, K., Nangle, L.A., Jang, J., Longo-Guess, C.M., Cook, S.A., Davisson, M.T, Sundberg, J.P., Schimmel, P., and Ackerman, S.L. (2006). Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* 443:50-55.

LéJohn, H. B., Cameron, L. E., Yang, B. and Rennie, S. L. (1994). Molecular characterization of an NAD-specific glutamate dehydrogenase gene inducible by L-glutamine: Antisense gene pair arrangement with L-glutamine-inducible heat shock70-like protein gene. *J. Biol. Chem.* 269: 4523-4531.

LéJohn, H. B., Cameron, L. E., Yang, B., MacBeath, G., Barker, D. S. and Willams, S. A. (1994). Cloning and analysis of a constitutive heat shock (cognate) protein 70 gene inducible by L-glutamine. *J. Biol. Chem.* 269: 4513-4522.

Link, A.J., Mock, M.L., Tirrell, D.A. (2003). Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* 14(6):603-609.

Liu, Y. and Kuhlman, B. (2006). RosettaDesign server for protein design. *Nucleic Acids Research* 34:235-238.

Lockless, S.W., Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286(5438):295-299.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., Butt T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Science* 15:182-9.

Michaelis, L., and Menten, M.L. (1913). Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift* 49:334-336.

Paley, E. L., Smelyanski, L., Malinovskil, V., Subbarayan, P. R., Berdichevsky, Y., Posternak, N., Gershoni, J. M., Sokolova, O., Denisova, G. (2007). Mapping and molecular characterization of novel monoclonal antibodies to conformational epitopes on NH(2) and COOH termini of mammalian tryptophanyl-tRNA synthetase reveal link of the epitopes to aggregation and Alzheimer's disease. *Mol. Immunol.* 44(4):541-557.

Park, S. G., Kim, S. (2006). Do Aminoacyl-tRNA synthetases have biological functions other than in protein biosynthesis? *IUBMB Life* 58(9):556-8.

Pham, Y., Li, L., Kim, A., Erdogan, O., Weinreb, V., Butterfoss, G.L., Kuhlman, B., Carter, C.W., Jr. (2007). A minimal TrpRS catalytic domain supports sense/antisense ancestry of class I and II aminoacyl-tRNA synthetases. *Molecular Cell* 25(6), 851-62.

Prætorius-Ibba, M., R. Rogers, et al. (2005). Association between archaeal prolyl- and leucyl-tRNA synthetases enhances tRNAPro aminoacylation. *J. Biol. Chem* 280(28): 26099–26104.

Praetorius-Ibba, M., Stange-Thomann, N., Kitabatake, M., Ali, K., Soll, I., Carter, C.W., Jr., Ibba, M., Soll, D. (2000). Ancient adaptation of the active site of tryptophanyl-tRNA synthetase for tryptophan binding. *Biochemistry* 39(43):13136-43.

Ribas de Pouplana, L. and Schimmel, P. (2001). Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. *Cell* 104: 191-193.

Rodin, A.S., Rodin, S.N. and Carter, C.W., Jr. (2009). On primordial sense-antisense coding. *J. Mol. Evol.* 69(5):555-567.

Rodin, S. N. and Ohno, S. (1995). Two Types of Aminoacyl-tRNA Synthetases Could be Originally Encoded by Complementary Strands of the Same Nucleic Acid. *Orig. Life Evol. Biosph.* 25: 565-589.

Sampath, P., Mazumder, B., Seshadri, V., Gerber, C., Chavatte, L., Kinter, M., Ting, S., Dignam, J. Kim, S., Driscoll, D. (2004). Noncanonical Function of Glutamyl-Prolyl-tRNA Synthetase Gene-Specific Silencing of Translation. *Cell* 119(2): 195-208.

Schimmel, P., and Schmidt, E. (1995). Making connections: RNA-dependent amino acid recognition. *Trends in Biochem. Sci.* 20(1):1-2.

Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., Baker, D. (2005). Progress in modeling of protein structures and interactions. *Science* 310(5748):638-42.

Schwob, E. and Soll, D. (1993). Selection of a minimal glutaminyl-tRNA synthetase and the evolution of class I synthetases. *EMBO J.* 12:5201-5208.

Seburn, K. L., Nangle, L. A., Cox, G. A., Schimmel, P., Burgess, R. W. (2006). An active dominant mutation of glycyl-tRNA synthetase causes neuropathy in a Charcot-Marie-Tooth 2D mouse model. *Neuron* 51(6):715-26.

Sievers, A., Beringer, M., Rodnina, M.V. and Wolfenden, R. (2004). The ribosome as an entropy trap. *Proc. Natl. Acad. Sci.* 101 7897-7901.

Starzyk, R. M., Burbaum, J.J., et al. (1989). Insertion of new sequences into the catalytic domain of an enzyme. *Biochemistry* 28(21): 8479-8484.

Traut, T. (2008). Regulatory allosteric enzymes (Book). Chapter 1:11-14.

Tzima, E., Reader, J.S., Irani-Tehrani, M., Ewalt, K.L., Schwartz, M.A., and Schimmel, P. (2005). VE-cadherin links tRNA synthetase cytokine to anti-angiogenic function. *J. Biol. Chem.* 280(4): 2405-2408.

Wakasugi, K., Slike, B.M., Hood, J., Otani, A., Ewalt, K.L., Friedlander, M. Cheresh, D.A., Schimmel, P. (2002). A human aminoacyl-tRNA synthetase as a regulator of angiogenesis. *Pro. Natl. Aca. Sci.* 99(1):173-177.

Williams, T.A., Wolfe, K.H. and Fares, M.A. (2009) No Rosetta Stone for a sense-antisense origin of aminoacyl-tRNA synthetase classes. *Mol. Biol. Evol.* 26(2):445-450.

Williamson, R.M. and Oxender, D.L. (1990). Sequence and structural similarities between the leucine-specific binding protein and leucyl-tRNA synthetase of Escherichia coli. *Proc. Natl. Acad. Sci*. 87:4561-4565.

Wolfson, W. D., Pleiss, J. A. and Uhlenbeck, O.C. (1998) A new assay for tRNA aminoacylation kinetics. *RNA* 4:1019-1023.

Yang, X.L., Otero, F.J., Ewalt, K.L., Liu J., Swairjo, M.A., Kohrer, C., RajBhandary, U.L., Skene, R.J., McRee, D.E., Schimmel, P. (2006). Two conformations of a crystalline human tRNA synthetase-tRNA complex: implications for protein synthesis. *EMBO J.* 25(12):2919-29.

Yang, X.L., Schimmel, P. and Ewalt, K.L. (2004). Relationship of two human tRNA synthetases used in cell signaling. *Trends in Biochem. Sci.* 29(5):250-256.

Yin, Y. and Carter, C.W., Jr. (1996). Incomplete factorial and response surface methods in experimental design:yield optimization of tRNA(Trp) from *in vitro* T7 RNA polymerase transcription. *Nucleic Acids Research* 24(7):1279-1286.

Zhang, Y., Baranov, P.V., Atkins, J.F. and Gladyshev, V.N. (2005). Pyrrolysine and Selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* 280:20470-751.

Zhou, X-L., Zhu, B. and Wang, E-D. (2008). The CP1 domain of leucyl-tRNA synthetase is crucial for amino acid activation and post-transfer editing. *J. Biol. Chem.* 283:36608-36616.