

**EVOLUTIONARY SYNTHETIC BIOLOGY: STRUCTURE/FUNCTION
RELATIONSHIPS WITHIN THE PROTEIN TRANSLATION SYSTEM**

A Thesis
Presented to
The Academic Faculty

by

Ercan Cacan

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Biology

Georgia Institute of Technology
December 2011

**EVOLUTIONARY SYNTHETIC BIOLOGY: STRUCTURE/FUNCTION
RELATIONSHIPS WITHIN THE PROTEIN TRANSLATION SYSTEM**

Approved by:

Dr. Eric A. Gaucher, Advisor
School of Biology
Georgia Institute of Technology

Dr. Roger M. Wartell
School of Biology
Georgia Institute of Technology

Dr. Brian K. Hammer
School of Biology
Georgia Institute of Technology

Date Approved: August 19th, 2011

I dedicate this work to my daughter Hazal Elif and my wife Hilal Hatun. This work would not have been possible without their perpetual love, faith, support, and encouragement.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Dr. Eric A. Gaucher for guiding, supporting and encouraging me throughout the graduate study. I would like to thank him for his unconditional availability and scientific inputs for every step in the research. Without his guidance and persistent help this research would not have been possible.

I owe a huge debt of gratitude to my family for their constant love and support that allow me to take every challenge with confidence. I like to thank my friends Sibel and Mustafa Kara.

I am also grateful to my lab-mates; Ryan Randall, Joshua Stern, Ziming Zhao, Dr. Megan Cole, and Dr. Betul Aslan for their collegiality, immeasurable support and technical guidance towards the research and for being such a huge family. I am also exceedingly grateful to my lab-mate and collaborator James Kratzer for his technical assistance. I would like to thank the departmental and university staff for rendering services cordially and efficiently.

I also would like to thank Dr. Matthew Hartman and his lab members for their collaboration.

Finally, I am grateful to my thesis committee members, Dr. Roger Wartell, Dr. Brian Hammer for their support, encouragement, feedback and suggestions that helped me in my research.

TABLE OF CONTENTS

	<u>Page #</u>
ACKNOWLEDGEMENTS	IV
LIST OF TABLES	IX
LIST OF FIGURES	X
LIST OF SYMBOLS AND ABBREVIATIONS	XII
SUMMARY	XIV
CHAPTER-I	
1. ENGINEERING tRNA FOR THE EFFICIENT INCORPORATION OF UNNATURAL AMINO ACIDS INTO THE PROTEIN TRANSLATION SYSTEM	1
1.1 BACKGROUND AND SIGNIFICANCE	1
Evolutionary Synthetic Biology	1
Overview Translation Mechanisms	2
Engineering Protein Synthesis System	4
Transfer RNA	4
Structure of Transfer RNA	9
G:U Non-Watson Crick Base Pair	10
Engineering Valine-tRNA	12
1.2. MATERIALS AND METHODS	14
<u>Materials</u>	14
Bacterial Strains: <i>Escherichia coli</i> strain and genotypes	14

Vector	14
Oligonucleotides	14
<u>Methods</u>	17
Molecular biology techniques	17
Primer design and template construction	17
QIAGEN Gel Extraction protocol	18
Cloning Procedures	19
Plasmid DNA Purification	20
Sequencing	21
Cleanup Amplified PCR Reaction: MinElute	21
<i>In Vitro</i> transcription	21
1.3. RESULTS AND DISCUSSIONS	25
Determine G:U Mutations in Valine tRNA	25
Cloning and Amplification of DNA	27
<i>In Vitro</i> Transcription	29
<i>In Vitro</i> Translation (PURE System)	31
Assay with Glutamate	32
Assay with Glutamine	34
1.4. CONCLUSION	36
1.5. APPENDIX	38
1.6. BIBLIOGRAPHY	39
CHAPTER II	
2. FUNCTIONAL DIVERGENCE AMONG HOMOLOGOUS PROTEINS	41
2.1 BACKGROUND AND SIGNIFICANCE	41

Functional Divergence	41
Rate heterogeneity and Heterotachy	41
Predicting Functional Divergence	48
Elongation Factor Tu	49
Elongation Factor 1A	51
2.2 MATERIALS AND METHODS	56
<u>Materials</u>	56
<i>Escherichia coli</i> strains and genotypes	56
Vectors	57
General Medias and Buffers	57
Genes and Primer	58
<u>Methods</u>	58
Preparation of Competent Cells	58
Subcloning EF-Ts and eEF1B	59
Recombinant PCR	59
QIAgen Gel Extraction protocol	60
Chemical Transformation	61
Analyzing Transformants	61
Plasmid DNA Purification	61
Sequencing	62
Expression and Purification of EF-Tu, EF-Ts, and eEF1B	62
Expression and Purification of EF-Tu variants, eEF1A, and its Variants	63
Refolding Denatured-Purified Protein	64

Removing His-tag from EF-Ts and eEF1B	64
Protein Quantification	65
Pull-Down Assay	65
Native Gel Electrophoresis	66
2.3 RESULTS AND DISCUSSIONS	67
Detecting Functionally Diverse Sequences	67
Detecting Mutated Sites	71
Subcloning EF-Ts and eEF1B	74
Expression and Purification of EF-Tu, EF-Ts, and eEF1B	75
Expression and Purification of EF-Tu variants, eEF1A, and its Variants under Denature Conditions	76
Refolding Purified Protein	77
Removing His-tag from EF-Ts and eEF1B	77
Interaction between EF-Tu:EF-Ts and eEF1A:eEF1B	78
Native Gel Analysis	78
Size-Exclusion Chromatography	80
Pull-Down Assay	82
Binding Ability of EF-Tu Variants to EF-Ts or eEF1B	83
Binding Ability of eEF1A Variants to EF-Ts or eEF1B	86
2.4. CONCLUSION	88
2.5. APPENDIX	90
2.6. BIBLIOGRAPHY	95

LIST OF TABLES

	<u>Page #</u>
Table1.1: Relationships between different tRNA species for EFTu with the presence of G:U base pairs in the Acceptor or T stem and their recruitment into the genetic code.	11
Table1.2: The primers that were used throughout the first chapter.	13
Table 1.3: Assembly of PCR reaction	17
Table 1.4: PCR cycling parameters.	17
Table 1.5: Assembly of TOPO cloning reaction for chemical transformation.	18
Table 1.6: Assembly of <i>in vitro</i> transcription reaction.	20
Table 2.1. General PCR cycling parameters for subcloning	57
Table 2.2: Considered statistical support and parameters to generate mutant variants.	69

LIST OF FIGURES

	<u>Page #</u>
Figure 1.1: The mechanism of protein synthesis	3
Figure 1.2: ATP-dependent of aminoacylation of tRNA	5
Figure 1.3: Thermodynamic compensation of amino acids and tRNAs to EF-Tu	7
Figure 1.4: Structure of tRNAs	8
Figure 1.5: G:U wobble base pair and electrostatic potential.	10
Figure1.6: Polyacrylamide Gel Electrophoresis of RNA.	22
Figure 1.7: Engineered Val-tRNA variants.	25
Figure 1.8: Gradient PCR.	26
Figure 1.9: Insertion of PCR product into Topo Vector.	27
Figure 1.10: Amplified DNA-tRNA.	28
Figure1.11: <i>In vitro</i> transcription	29
Figure 1.12: Quantify <i>in vitro</i> transcription product.	30
Figure 1.13: Assay with glutamate.	32
Figure 1.14: Assay with Glutamine.	34
Figure 1.15: G:U base pair distribution on tRNA based on affinity to EF-Tu	36
Figure 2.1. A hypothetical scheme to show the rate heterogeneity of sites in a given gene	40
Figure 2.2. Homotachous and heterotachous rate heterogeneity models	41
Figure 2.3. Distribution of heterotachy sites onto the three dimensional structure of EF-Tu	43
Figure 2.4. An Example of heterotachy at a particular site in Elongation factor gene family	45
Figure 2.5. Diagrams of site-specific functional divergence, type-0, type-I and type-II amino acid configurations.	46

Figure 2.6. Structural organization of EF-Tu and eEF1A.	48
Figure 2.7. Comparison of sequence similarity among EF-Tu from <i>E. coli</i> and eEF1A from <i>S. cerevisiae</i> .	49
Figure 2.8. The tree used as input for DIVERGE for selecting the bacterial and eukaryotic clusters	52
Figure 2.9. Sequence alignment of Eukaryotic and Bacterial EFs	66
Figure 2.10a. Sequence alignment of eEF1A variants.	71
Figure 2.10b. Sequence alignment of EF-Tu variants	72
Figure 2.11. SDS-PAGE gel analysis of purified-recombinant EF-Tu, EF-Ts, and eEF1B proteins.	73
Figure 2.12. SDS-PAGE gel analysis of thrombin cleavage of EF-Ts. Purified N-terminal 6XHis-tag EF-Ts cleaved by thrombin.	76
Figure 2.13. Native gel analysis of interactions EF-Tu and EF-Ts.	78
Figure 2.14. Interaction between <i>E. coli</i> EF-Tu:EF-Ts complexes analyzed by HiPrep Superdex-200 size-exclusion column chromatography.	79
Figure 2.15. Interaction between <i>Saccharomyces cerevisiae</i> eEF1A:eEF1B complexes analyzed by HiPrep Sephacryl S-200 size-exclusion column chromatography	80
Figure 2.16. SDS-PAGE analysis of interaction between EF-Tu:EF-Ts and eEF1A:eEF1B.	81
Figure 2.17. SDS-PAGE gel analysis of binding ability of nucleotide exchange factors to EF-Tu variants.	83
Figure 2.18. SDS-PAGE gel analysis of binding ability of nucleotide exchange factors to eEF-1A variants.	85

LIST OF SYMBOLS AND ABBREVIATIONS

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
mRNA	Messenger RNA
tRNA	Transfer RNA
rRNA	Ribosomal RNA
GTP	Guanosine triphosphate
GDP	Guanosine diphosphate
EF-Tu	Elongation Factor-Tu
EF-Ts	Bacterial nucleotide exchange factor-Ts
eEF1A	Eukaryotic Elongation Factor-1A
eEF1B	Eukaryotic nucleotide exchange factor-1B
EF	Elongation Factor
aa-tRNA	Aminoacyl-tRNAs
aaRS	Aminoacyl synthesis
ATP	Adenosine triphosphate
AMP	Adenosine monophosphate
Glu-tRNA	Glutamate tRNA
G:U	Guanidine:Uracil
G:C	Guanidine:Cytosine
A:U	Adenine:Uracil

Val-tRNA	Valine-tRNA
Glu	Glutamate
Gln	Glutamine
IDT	Integrated DNA Technologies Inc.
PCR	Polymerase chain reaction
LB	Luria-Bertani
SDS	Sodium Dodecyl sulfate
PAGE	Polyacrylamide Gel Electrophoresis
PURE	Protein synthesis Using Recombinant Elements
EtBr	Ethidium Bromide
IPTG	Isopropyl β -D-1-thiogalactopyranoside
MWCO	Molecular weight cut-off
TCA	Trichloroacetic acid

SUMMARY

Production of mutant biological molecules for understanding biological principles or as therapeutic agents has gained considerable interest recently. Synthetic genes are today being widely used for production of such molecules due to the substantial decrease in the costs associated with gene synthesis technology. Along one such line, we have engineered tRNA genes in order to dissect the effects of G:U base-pairs on the accuracy of the protein translation machinery. Our results provide greater detail into the thermodynamic interactions between tRNA molecules and an Elongation Factor protein (termed EF-Tu in bacteria and eEF1A in eukaryotes) and how these interactions influence the delivery of aminoacylated tRNAs to the ribosome. We anticipate that our studies not only shed light on the basic mechanisms of molecular machines but may also help us to develop therapeutic or novel proteins that contain unnatural amino acids. Further, the manipulation of the translation machinery holds promise for the development of new methods to understand the origins of life.

Along another line, we have used the power of synthetic biology to experimentally validate an evolutionary model. We exploited the functional diversity contained within the EF-Tu/eEF1A gene family to experimentally validate the model of evolution termed ‘heterotachy’. Heterotachy refers to a switch in a site’s mutational rate class. For instance, a site in a protein sequence may be invariant across all bacterial homologs while that same site may be highly variable across eukaryotic homologs. Such patterns imply that the selective constraints acting on this site differs between bacteria and eukaryotes. Despite intense efforts and large interest in understanding these patterns, no studies have experimentally validated these concepts until now. In the present study, we analyzed EF-Tu/eEF1A gene family members between bacteria and eukaryotes to identify heterotachous patterns (also called Type-I functional divergence). We

applied statistical tests to identify sites possibly responsible for biomolecular functional divergence between EF-Tu and eEF1A. We then synthesized protein variants in the laboratory to validate our computational predictions. The results demonstrate for the first time that the identification of heterotachous sites can be specifically implicated in functional divergence among homologous proteins.

In total, this work supports an evolutionary synthetic biology paradigm that in one direction uses synthetic molecules to better understand the mechanisms and constraints governing biomolecular behavior while in another direction uses principles of molecular sequence evolution to generate novel biomolecules that have utility for industry and/or biomedicine.

CHAPTER I

1. ENGINEERING tRNA FOR THE EFFICIENT INCORPORATION OF UNNATURAL AMINO ACIDS INTO THE PROTEIN TRANSLATION SYSTEM

1.1. BACKGROUND AND SIGNIFICANCE

Evolutionary Synthetic Biology

Synthetic biology is a new area of biological research and has received considerable interest in the last decade. This field involves the synthesis of novel biological systems that are not usually found in nature. Synthetic biologists attempt to assemble unnatural components to generate novel systems with the goal of better understanding the principles of biology and evolution. The term synthetic biology was first associated with research into the engineering of control elements in existing genomes and the build up modified genomes by the Polish geneticist Wacław Szybalski in 1974. In 1980, Barbara Hobom used the term to describe bacteria that had been genetically engineered using recombinant DNA technology (Hobom, B. 1980). In 2000, Eric Kool described the term as the synthesis of unnatural organic molecules that function in living systems (Rawls, R. 2000). Consequently, most of synthetic biologists are attempting to develop tools and methods to control interactions between DNA, RNA, and proteins because these interactions lie at the heart of living systems.

Evolutionary synthetic biology aims to exploit evolutionary principles and models to develop engineered molecules and organisms. It also proposes to generate synthetic biomolecules to improve our understanding of evolutionary principles and structure/function relationships of biomolecules at the sequence level (Gaucher, E. Suddath Symposium, 2011).

Overview Translation Mechanisms

The majority of biological activities are carried out by proteins. The linear order of amino acids in each peptide directly (and sometimes indirectly) determines the structure and activity of a protein. Therefore, assembly of amino acids in their correct order is crucial to produce functional proteins in organisms. Translation is a process that uses genetic information to create biological function. Translation takes place on the ribosome where mRNA functions as a template to join amino acids carried by an elongation factor protein (EF-Tu in bacteria and eEF1A in eukaryotes) and specific aminoacyl-tRNAs (aa-tRNA), to produce a polypeptide chain. The basic mechanism for translation is given in Figure 1.1. Protein synthesis is generally divided into three fundamental phases, which are Initiation, Elongation, and Termination. During initiation of protein synthesis, the large and small ribosomal subunits assemble after mRNA associates with the small subunit of the ribosome. Amino acids are attached to their corresponding tRNAs by their cognate aminoacyl tRNA synthesis (aaRS). The resulting complex is called aa-tRNA. After activation of amino acid, the aa-tRNA complex is carried to the A site of the ribosome by EF-Tu. Correct codon and anticodon pairing results in the release of EF-Tu from the aa-tRNA complex. The target amino acid is transferred to the growing polypeptide chain. After each successive addition of amino acid, the de-acylated tRNA moves through the ribosome and is released at the E-site of the ribosome. This is known as the elongation phase. Translation keeps going in this manner until whole genetic information in mRNA is translated into protein. When the ribosome encounters a stop codon, protein synthesis enters the termination phase where the protein synthesis is ended by termination factors and ribosomal subunits separate.

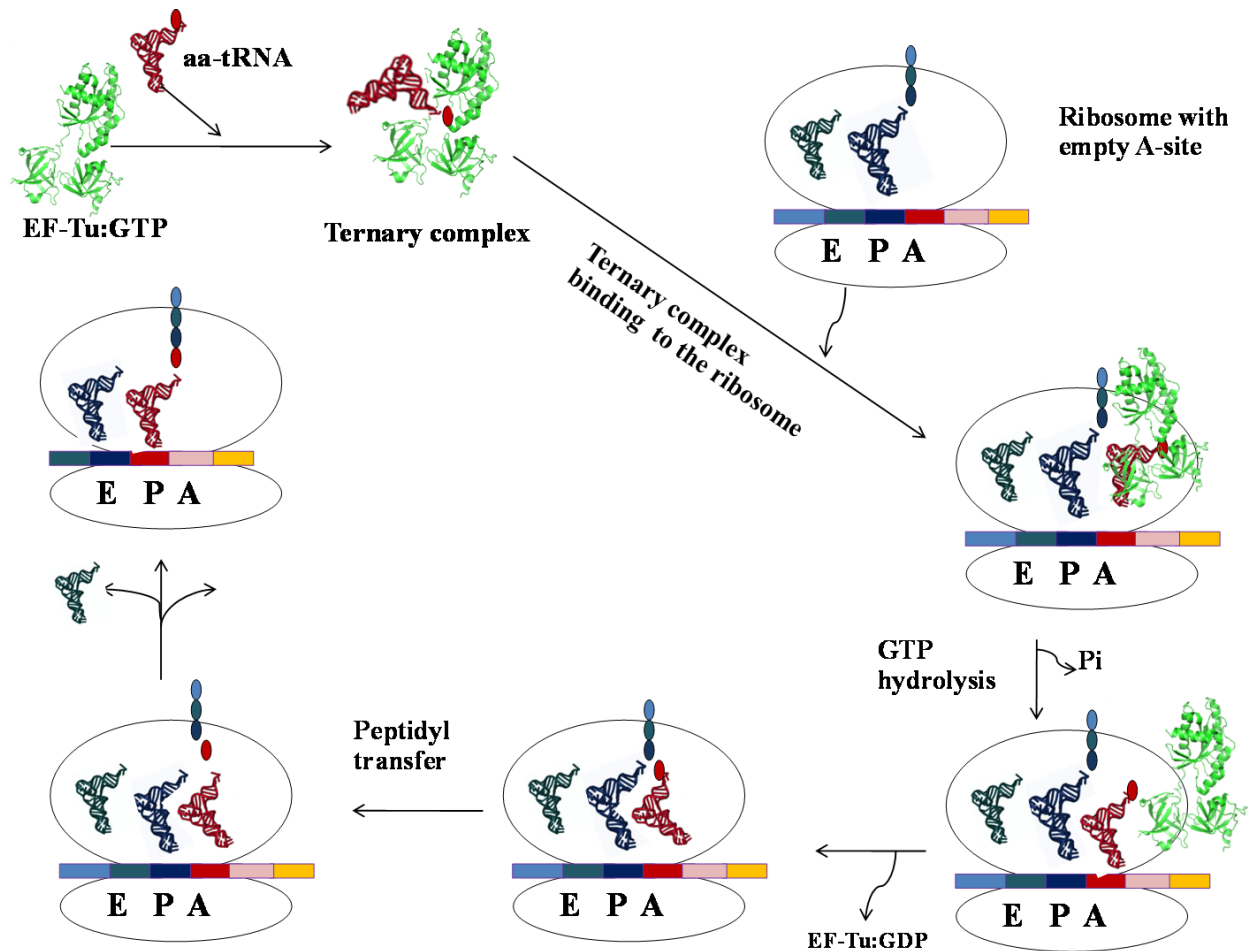


Figure 1.1. The general mechanisms of protein synthesis. Each amino acid is recognized by its specific aminoacyl-tRNA synthetase (aaRS) and charged with the appropriate tRNA to form an aminoacyl-tRNA complex (aa-tRNA). aa-tRNA binds to the Bacterial Elongation Factor Tu (EF-Tu:GTP) to form a ternary complex, which binds to the ribosome. Once the correct codon and anticodon pair occurs between mRNA and tRNA, GTP is hydrolyzed to GDP. This hydrolysis cause conformational change which results in the disassociation of EF-Tu from aa-tRNA and the aa-tRNA then moves to the E site of the ribosome.

In all steps, the required energy is provided by ATP or GTP hydrolysis. In general, the fidelity of protein synthesis in bacteria depends of the accuracy of codon-anticodon base-pairing, correct charging of amino acid onto cognate tRNA, and EF-Tu's ability to discriminate between correctly charged and incorrectly charged tRNAs (Yarus *et al.*, 1995; Ibba *et al.*, 1997; LaRiviere *et al.*, 2001).

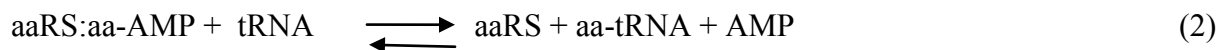
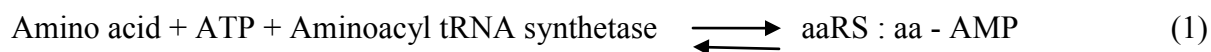
Engineering protein synthesis system

In the past several years, there has been a great deal of interest in engineering RNA and proteins with the intentions that such engineered components may lead to both a better understanding of protein translation and help produce novel molecules that have therapeutic value. Despite this interest, manipulating the protein translation system is not straightforward. For instance, it is not easy to incorporate just any amino acid into the translation machinery because there are many checks-and-balances during protein synthesis, such as aminoacylation of tRNA, delivery of aa-tRNAs by EF-Tu, and incorporation of an amino acid (natural or unnatural) into a peptide chain by the ribosome. It is not exactly clear how the translation machinery achieves all of this specificity. Therefore, we have attempted to manipulate specific interactions in the protein translation machinery to better understand specificity. For instance, we are focused on the interactions between tRNA and EF-Tu with the intentions that we can manipulate these interactions to help us incorporate unnatural amino acids during protein synthesis.

Transfer RNA (tRNA)

The genetic code consists of 64 possible triplet codons which can be translated into a polypeptide composed of 20 possible amino acids. The polypeptides are coded by the triplet

codons on the mRNA (Crick, F. H., *et al.* 1961). The role of recognizing the triplet codons is given to the tRNA which is a relatively small RNA molecule, 72 to 95 nucleotides in length, and present in all three domains of life. There are about 46 different tRNAs present in *E.coli* and they are necessary for translating the genetic code (Sprinzl, Mathias. 2006). These tRNAs have two crucial functions: to recognize a specific codon on the mRNA and to accept the coded amino acid for transport to the A-site of the ribosome. tRNA identity underlies its capacity to be specifically catalyzed by aaRSs and thus to be responsible for the correct translation of DNA into the protein (RajBhandary UL. 1994). The process of catalyzing an amino acid with its cognate tRNA is called aminoacylation (Figure 1.2.). Fundamentally, aminoacylation reaction is carrying out in two steps:



In the first step, the amino acid is activated with ATP. In the second step, activated-amino acid is transferred to the 3' end of the tRNA by aaRS.

The aminoacylation reaction must be quite specific for both their cognate tRNA and amino acid. Although each aaRS is highly specific for its cognate amino acid, in some cases, it is hard to distinguish them due to high similarity between the functional groups for certain amino acid side chains.

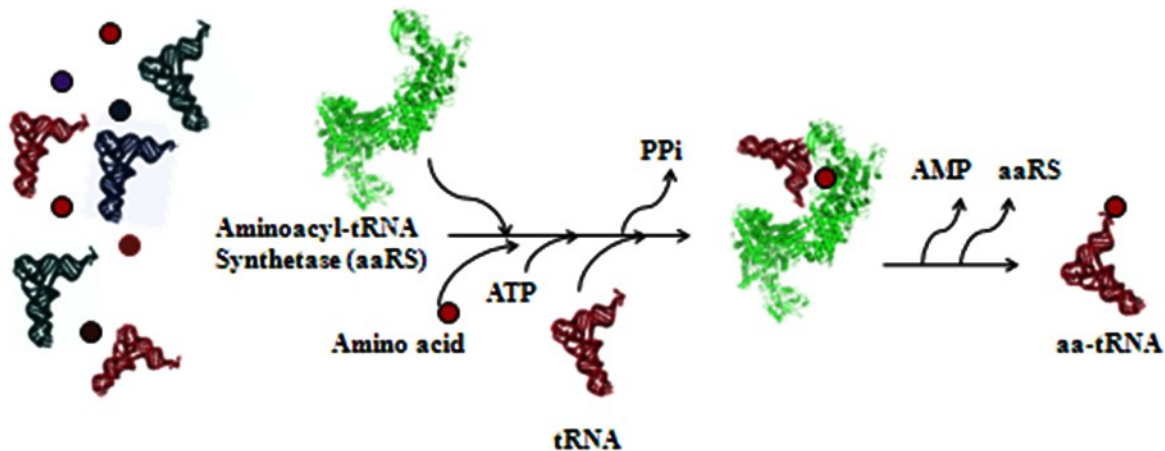


Figure 1.2. ATP-dependent aminoacylation of tRNA. An uncharged tRNA and appropriate (cognate) amino acid is aminoacylated by a specific aminoacyl-tRNA synthetase in the presence of ATP. ATP is hydrolyzed to AMP and once the amino acid is charged onto the tRNA, aminoacyl-tRNA synthetase disassociates from complex and aminoacyl-tRNA is now ready to participate in protein synthesis.

During aminoacylation, each amino acid is accepted by its cognate aaRS. This part of translation is known as the specific phase because the aminoacylation reaction is very specific for both the tRNA and the amino acid (Dale and Uhlebeck, 2005). The traditional view of tRNAs was that they were interchangeable components and passive adaptors of protein synthesis (Woese, C.R. 2001). However, Uhlenbeck *et al.* have demonstrated that each tRNA sequence has co-evolved with its cognate amino acid so tRNAs are active molecules and adapted to meet specific kinetic requirements of translation, so called *thermodynamic compensation*. Specifically, this means that the 20 cognate tRNAs:aa have similar thermodynamic contributions in their binding to EF-Tu. However, tRNA itself has a wide spectrum of thermodynamic contributions to EF-Tu binding. Uhlenbeck *et al.* essentially mischarged specific tRNAs with several different amino acids. They examined the interactions between those mischarged-tRNAs and their ability to bind EF-Tu (Schrader *et al.*, 2009). They determined whether EF-Tu recognizes a specific amino acid and binds tightly to it or recognizes a specific tRNA and binds tightly to it. This

relationship is inversely proportional (Figure 1.3). For example, if you have Glutamate – tRNA (Glu-tRNA) and that tRNA interacts strongly with EF-Tu, then the Glutamate itself will not interact strongly with EF-Tu. Conversely, if you have an amino acid that interacts strongly (e.g. glutamine) with EF-Tu, then the cognate tRNA will not interact strongly with EF-Tu. Thus, if you take a weak amino acid and aminoacylate it onto a weak tRNA then presumably that mischarged tRNA will not bind strongly enough to EF-Tu to be delivered to the ribosome. Therefore, if we want to make a protein or peptide that contains unnatural amino acids (ones that presumably are not able to strongly interact with EF-Tu) then we should attempt to bind such unnatural amino acids to a strong tRNA. In this situation, the complex should be delivered to the ribosome. So if we put this concept to work, we may convert a weak binding tRNA into a strong binding tRNA in terms of its interactions with EF-Tu and this presumably allows us then to attach a variety of unnatural amino acids onto modified tRNAs.

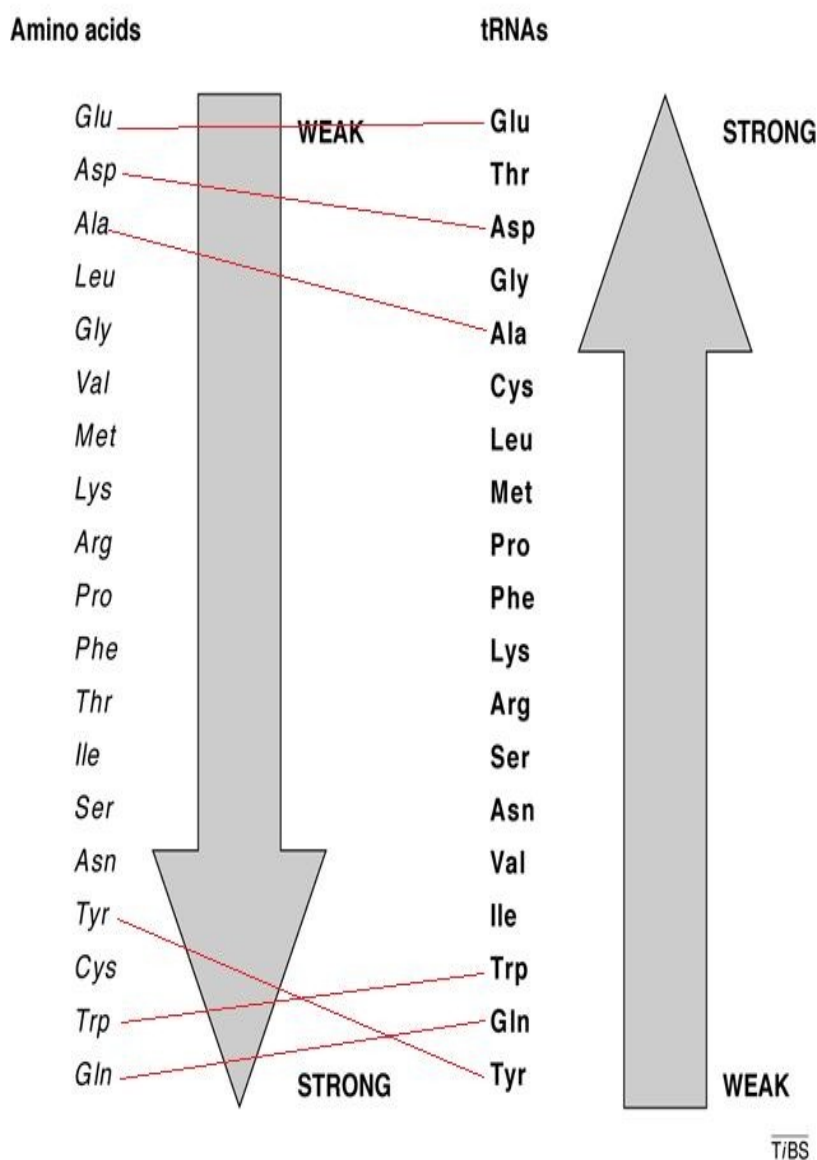


Figure 1.3. Thermodynamic compensation of amino acids and tRNAs to EF-Tu. The order of thermodynamic contributions for most amino acids binding to EF-Tu has been determined experimentally and some of them have been predicted. The amino acid and tRNA hierarchies are approximately inversely proportional because the thermodynamic contributions of the amino acid and the tRNA balance one another. For example the Glu-tRNA is the strongest binder among the tRNAs, and the amino acid glutamate is the weakest binder to EF-Tu among the amino acids. Due to thermodynamic compensation, Glu-tRNA^{Glu} and Gln-tRNA^{Gln} show similar affinities towards their bind to EF-Tu (Dale and Uhlenbeck, 2004-2005).

Structure of tRNA

Specialized tRNA functions in protein synthesis are mostly related to typical structural features. Given the important roles of tRNAs in living systems, it is desirable to manipulate tRNAs for understanding tRNA structure-function relationship and consequently generating proteins and organisms with new properties. This will enable us to provide novel tools for understanding biology in molecular terms and in studying various biological questions. **A**

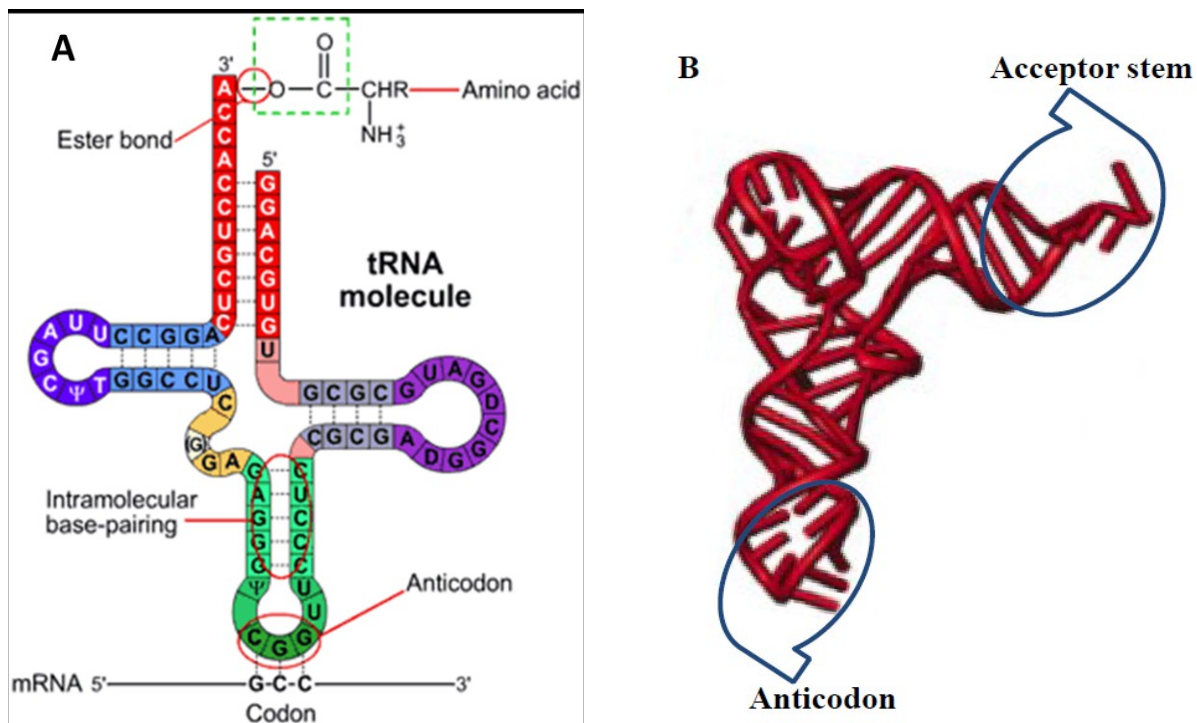


Figure 1.4. Structure of tRNAs. A) Cloverleaf folding highlights the different domains of the tRNA molecule. Length of tRNA molecule ranges from 72-95 nucleotides, with an average of 76 nucleotides. Structural differences mainly originate from length variation in the D-arm and variable region. The CCA sequence at the 3' end is found in nearly all tRNAs. Attachment of an amino acid to the 3' A yields an aminoacyl-tRNA. B) Three-dimensional model of tRNA. L shaped folding showing how the tRNA domains organize themselves.

Two-dimensional tRNA structure is usually described in cloverleaf-model. All tRNAs contain the residues CCA at the 3' terminus. The cloverleaf-modeled tRNA structure has four arms: (a) The acceptor stem (A) accepts the amino acid at the CCA-3' sequence and binds to the EF-Tu molecule; (b) The T ψ C (T) loop is so named because of the presence of this triplet sequence. This loop also binds to the EF-Tu molecule; (c) The anticodon loop located at the opposite end of the acceptor arm. This arm carries the anticodon triplet in the center of the loop and is responsible for codon-anticodon base pairing. (d) The D loop is so named because it is rich in dihydrouridine, a modified base. There is one extra arm located between the T ψ C and anticodon arm of variable length. The significant function of this arm has not been determined yet for most tRNAs. One of the structural distinctive features of tRNA is that they contain nonstandard bases and pairings. The most significant such pairing is the G:U non Watson-Crick base-pair.

G:U Non-Watson Crick Base Pair

The G:U wobble base pair is a non-Watson-Crick base pairing, which is essential for RNA structure. G and U bases interact via two hydrogen bonds in the same manner that Watson-Crick base pairing occurs (Ladner *et al.*, 1975). Subsequent studies have shown that the G:U pair provides unique recognition sites and can have a major impact on how ancillary binding interactions occur (Gabriel *et al.*, 1996; Varani *et al.*, 2000; Agris *et al.*, 2007). Studies have also showed that tRNA binds to EF-Tu via its Acceptor and T stems where G:U wobble base pairs may be responsible for either weak- or strong-binding of this complex. Again, we have learned from previous studies that the G:U wobble base-pair presents an electronegative environment in the tRNA structure where EF-Tu binds (Figure 1.5). Xu and his coworkers have found that the

GU pair increases negativity when compared with standard G:C or A:U base pairs in the major groove of RNA. Conversely, when in the minor groove, G:U decreases negativity. We have hypothesized that positive amino acids in EF-Tu should bind tightly to the G:U wobble base pair while negative amino acids bind weakly to the tRNA's region when interacting with G:U pairs. Our bioinformatic analyses have discovered that so-called strong binding tRNAs have a G:U pairing in particular regions of the tRNA structure while the so-called weak binding tRNAs have G:U in different regions of the tRNA structure. Typically, a G:U around the acceptor stem seems to be correlated to the strong binding tRNAs while the weak binding tRNAs mostly have G:U base-pairs around the T stem (Table 1.1).

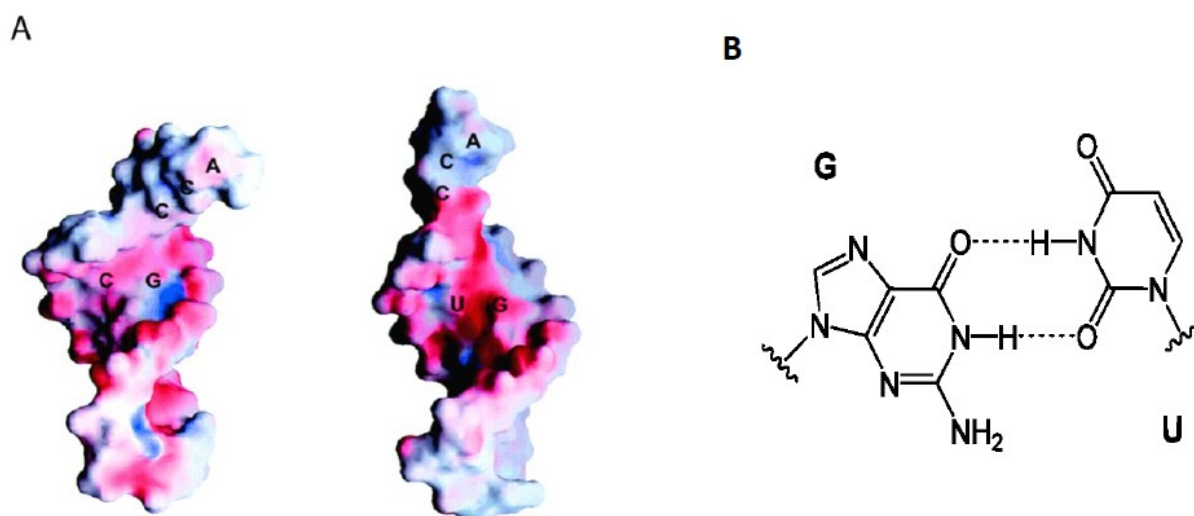


Figure 1.5. G:U wobble base pair. A) Surface electrostatic potential maps of Watson-Crick G:C and G:U wobble base pair. Red color represents negative potential and blue represents positive potential regions. B) Representative scheme between Guanine and Uracil, their interaction via two hydrogen bonds. Varani *et al.*, (2000).

Despite substantial interest and a few studies, no study has ever established a rule based on the distribution of the G:U wobble base-pair in tRNA structures. Therefore, we have attempted to determine sequence structure-function relationships of tRNAs and their interactions

with EF-Tu, regarding the G:U wobble pairing. To test our idea, we have elected to use Valine (Val)-tRNA as a model.

Table 1.1 Relationships between different tRNA species for EFTu with the presence of G:U base pairs in the Acceptor or T stem and their recruitment into the genetic code. All tRNA species come from *E. coli* K-12 strain.

tRNA	Anticodon	G:U base pair	Affinity for EFTu
Aspartate	GUC	49:65	strong
		4:69	
Glycine	GCC	49:65	strong
	CCC		
Threonine	CGU	49:65	strong
Alanine	GGC	3:70	strong
	UGC		
Phenylalanine	GAA	51:63	weak
Valine	UAC	50:64	weak
	GAC	49:65	studied here
f-Methionine	CAU	50:64	weak
Tryptophan	CCA	50:64	weak
Histidine	GUG	49:65	unknown
Isoleucine	GAU	5-68	
		49:65	

Engineering Valine-tRNA

We engineered Val-tRNA by introducing site specific mutations based on our bioinformatic analysis. For the present study, we used Val-tRNA because it shows moderate

binding to EF-Tu, which enable us to convert this tRNA into either a strong- or weak-binding tRNA. As a first step, we wanted to introduce individual mutations to the specific regions of the tRNA that interact with EF-Tu. Therefore, we introduced mutations to the acceptor and T stem of tRNA in order to create G:U wobble pairs. Nevertheless, to see the affect of G:U pairings at specific sites, we also introduced a mutation to eliminate the G:U pair located in wild type Val-tRNA at 50:64 position. To investigate our idea, we charged Val-tRNA with Glutamate (Glu) which is a weak amino acid and Glutamine (Gln) which is known a strong amino acid. We hypothesized that the Val-tRNA^{Gln} pair would participate in translation more efficiently than Val-tRNA^{Glu} because of the thermodynamic compensation to EF-Tu.

1.2. MATERIALS AND METHODS

Materials

Bacterial Strains: *Escherichia coli* strain and genotypes

For general cloning and screening, we used One Shot TOP10 Chemically Competent *E. coli* cells (*Invitrogen*) which provides a one-step cloning strategy for the direct insertion of *Taq* polymerase-amplified PCR products into a plasmid vector.

The genotype of TOP10 F⁻ *mcrA* Δ (*mrr-hsdRMS-mcrBC*) Φ 80*lacZ* Δ M15 Δ *lacX74* *recA1* *araD139* Δ (*ara-leu*) 7697 *galU* *galK* *rpsL* (StrR) *endA1* *nupG*

Vector

We used pCR2.1-TOPO vector ((*Invitrogen*) for chemical transformation of PCR product. **Antibiotic resistance:** carbenicillin and kanamycin. For details see appendix.

Oligonucleotides

Synthetic PCR primers were purchased from Integrated DNA Technologies (IDT) Inc. All primers used in this study are shown in Table 1.2. The primers were diluted up to 100 uM and stored at -20 °C.

Table 1.2: The primers used throughout this study. DNA sequences are also provided for each variant. Synthesized primers were used to construct template DNA and amplifying primers were used to amplifying the template.

Wild Type	DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GGC GGT TCG ATC CCG TCA TCA CCC ACC A- 3' Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GGC GGT T- 3'
-----------	--

	<p>Synthesize wt tRNA rev: 5'-TGG TGG GTG ATG ACG GGA TCG <u>AAC CGC CGA CCC CCT CC</u>- 3'</p> <p>Amplifying forward Primer: 5'-<u>TAA TAC GAC TCA CTA TAG</u> GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGG GTG ATG AC- 3'</p>
64 T→C	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GGC GGT TCG ATC CCG CCA TCA CCC ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC <u>AAG GAG GGG GTC GGC GGT T</u>- 3'</p> <p>Synthesize wt tRNA rev: 5'-TGG TGG GTG ATG GCG GGA TCG <u>AAC CGC CGA CCC CCT CC</u>- 3'</p> <p>Amplifying forward Primer: 5'-<u>TAA TAC GAC TCA CTA TAG</u> GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGG GTG ATG GC- 3'</p>
50 G→A	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT TCG ATC CCG TCA TCA CCC ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC <u>AAG GAG GGG GTC GAC GGT T</u>- 3'</p> <p>Synthesize wt tRNA rev: 5'-TGG TGG GTG ATG ACG GGA TCG <u>AAC CGT CGA CCC CCT CC</u>- 3'</p> <p>Amplifying forward Primer: 5'-<u>TAA TAC GAC TCA CTA TAG</u> GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGG GTG ATG AC- 3'</p>
50 G→A 65 C→T	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT TCG ATC CCG TTA TCA CCC ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC <u>AAG GAG GGG GTC GAC GGT T</u>- 3'</p> <p>Synthesize wt tRNA rev: 5'-TGG TGG GTG ATA ACG GGA TCG <u>AAC CGT CGA CCC CCT CC</u>- 3'</p> <p>Amplifying forward Primer: 5'-<u>TAA TAC GAC TCA CTA TAG</u> GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGG GTG ATA AC- 3'</p>
50 G→A 51 C→T	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAT GGT TCG ATC CCG TCA TCA CCC ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC <u>AAG GAG GGG GTC GAT GGT T</u>- 3'</p> <p>Synthesize wt tRNA rev: 5'-TGG TGG GTG ATG ACG GGA TCG <u>AAC CAT CGA CCC CCT CC</u>- 3'</p> <p>Amplifying forward Primer: 5'-<u>TAA TAC GAC TCA CTA TAG</u> GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGG GTG ATG AC- 3'</p>

50 G→A 72 C→T	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT TCG ATC CCG TCA TCA CCT ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT T- 3'</p> <p>Synthesize wt tRNA rev: 5'-TGG TAG GTG ATG ACG GGA TCG AAC CGT CGA CCC CCT CC- 3'</p> <p>Amplifying forward Primer: 5'-TAA TAC GAC TCA CTA TAG GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TAG GTG ATG AC- 3'</p>
50 G→A 71 C→T	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT TCG ATC CCG TCA TCA CTC ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT T- 3'</p> <p>Synthesize wt tRNA rev: 5'-TGG TGA GTG ATG ACG GGA TCG AAC CGT CGA CCC CCT CC- 3'</p> <p>Amplifying forward Primer: 5'-TAA TAC GAC TCA CTA TAG GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGA GTG ATG AC- 3'</p>
50 G→A 70 C→T	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT TCG ATC CCG TCA TCA TCC ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT T- 3'</p> <p>Synthesize wt tRNA rev: 5'-TGG TGG ATG ATG ACG GGA TCG AAC CGT CGA CCC CCT CC- 3'</p> <p>Amplifying forward Primer: 5'-TAA TAC GAC TCA CTA TAG GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGG ATG ATG AC- 3'</p>
50 G→A 69 A→T	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT TCG ATC CCG TCA TCT CCC ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GAC GGT T- 3'</p> <p>Synthesize wt tRNA rev: 5'-TGG TGG GAG ATG ACG GGA TCG AAC CGT CGA CCC CCT CC- 3'</p> <p>Amplifying forward Primer: 5'-TAA TAC GAC TCA CTA TAG GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGG GAG ATG AC- 3'</p>
70 C→T	<p>DNA sequence: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GGC GGT TCG ATC CCG TCA TCA TCC ACC A- 3'</p> <p>Synthesize wt tRNA fwd: 5'-GGG TGA TTA GCT CAG CTG GGA GAG CAC CTC CCT TAC AAG GAG GGG GTC GGC GGT T- 3'</p>

	<p>Synthesize wt tRNA rev: 5'-TGG TGG ATG ATG ACG GGA TCG <u>AAC CGC CGA CCC CCT CC</u>- 3'</p> <p>Amplifying forward Primer: 5'-<u>TAA TAC GAC TCA CTA TAG</u> G GGT GAT TAG CTC AGC- 3'</p> <p>Amplifying reverse Primer: 5'- TGG TGG ATG ATG AC- 3'</p>
--	---

Methods

Molecular biology techniques

Standard molecular biology protocols were used for DNA electrophoresis, restriction digestion, ligation, and bacterial transformation. Enzymes were purchased from *New England Biolabs* and *Promega*.

Primer design and template construction

For each variant we have two primers to construct the template and two primers to amplify template DNA. All primers were purchased from IDT. To synthesize the wild-type and variant genes, three PCR steps were applied; gradient, template construction-amplification, template amplification after sequence verification. Gradient PCR was performed to find optimum annealing temperatures for the variants. Template primers were mixed with 5X GoTaq buffer, dNTPs, GoTaq polymerase and water was added up to a 25 µL reaction volume. PCR was run, but after 4 cycles, amplifying primers were then added. The PCR conditions are summarized in Table 1.3 and 1.4.

Table 1.3: PCR reaction. The following PCR reaction set up for 25 μ L.

Component	Final Concentration	Volume/25 μ L rxn
Water	—	14.25 μ L
5X Go Taq colorless Buffer	5X	5 μ L
dNTPs	10mM	0.5 μ L
T7 Amplify-tRNA fwd primer	10 μ M	2 μ L
Amplify-tRNA rev primer	10 μ M	2 μ L
Synthesize tRNA fwd primer	1 μ M	1 μ L
Synthesize tRNA rev primer	1 μ M	1 μ L
Go Taq polymerase	5U/ μ L	0.25 μ L

Table 1.4: PCR cycling parameters. The annealing temperature varies depending on the variant's primers and template.

Step	Time	Temperature	Cycles
Initial Denaturation	3 min	95 °C	1X
Denaturation	30 sec	95 °C	
Annealing	30 sec	52 °C	34X
Extension	30 sec	72 °C	
Final Extension	6 min	72 °C	1X

QIAGEN Gel Extraction protocol

Fragments of DNA generated by the PCR reaction were separated using standard DNA electrophoresis (0.8% agarose gel). DNA bands corresponding to the desired products were identified using a UV transilluminator and bands were excised from Ethidium Bromide (EtBr) stained gels by using a scalpel. Separation of DNA from gel was achieved by using the QIAGEN Gel Extraction Kit and protocols supplied by the manufacturer (*QIAGEN*). Basically, the gel slice

was weighted in a 1.5 mL microcentrifuge tube and mixed with 3 volumes of QG buffer containing guanidine thiocyanate. Samples were incubated at 48 °C until the gel slice has completely dissolved and the optimal pH for DNA binding to the spin column was adjusted by using 3 M sodium acetate (pH 5.0). To increase the yield of DNA fragments, 1 gel volume of isopropanol was added to the sample and mixed. The sample was loaded into the column, provided by manufacturer, and centrifuged. The column was washed with PE buffer containing ethanol and the DNA was eluted with 50 µL sterile molecular grade water (pH 7.5) (*Mediatech Inc*). To increase DNA concentration, the tube was incubated for 3-5 min at room temperature before centrifugation. The purified DNA was stored at -20 °C.

Cloning Procedures

One Shot TOP10 Chemically Competent *E. coli* cells were used for cloning. Although a detailed-procedure can be obtained from manufacturer (*Invitrogen*), the basic protocol is as follows: cleaned-up PCR product was mixed with TOPO vector and salt solution. We added salt (200 mM NaCl; 10 mM MgCl₂) to increase transformation efficiency up to 3 fold. The mixture was incubated at room temperature for several minutes (min), and then the reaction was transformed to *E. coli* competent cells and incubated on ice for several min. After incubation, the cells were heat-shocked for 45 seconds at 42 °C and immediately transferred to ice. The sample was shaken at 37 °C for 1 hr in S.O.C. medium, and then plated on carbenicillin and kanamycin selective plates.

Next day, 5 single white colonies were taken from the selective plate and inoculated individually in 4 mL LB (Luria-Bertani) medium containing 100 µg/mL carbenicillin and kanamycin. The culture was placed in 37 °C incubator with 250 rpm shaking overnight.

Table 1.5: Set up TOPO cloning reaction for eventual transformation into chemically competent TOP10 One Shot *E. coli*.

Reagent	Chemically Competent <i>E. coli</i>
Fresh PCR product	0.5 to 4 μ L
Salt Solution	1 μ L
TOPO® vector	1 μ L
Water	add to a total volume of 5 μ L
Final Volume	6 μL

Plasmid DNA Purification

The plasmid DNA was isolated using QIAprep Spin Miniprep Kit from *QIAGEN*. The bacterial cells were transfer into a microcentrifuge tube and collected by centrifugation* at room temperature. The pellet resuspended in buffer P1 (lysis buffer and *RNase-A*), buffer P2 (contains sodium hydroxide), and buffer N3 (contains guanidine hydrochloride, acetic acid), respectively. After cells lysate was clarified by centrifugation, the supernatant was load on the QIAprep spin column and washed with Buffer PE. The plasmid DNA elution was achieved in a 1.5 mL sterile microcentrifuge tube by 50 μ L sterile molecular grade water (PH 7.5), *Mediatech, Inc.* The concentration of the samples was determined using Nanodrop 1000 spectrophotometer, *Thermo Scientific Inc.*

*All centrifugation steps were carried out at 13,000 rpm (~17,900 x g) in a conventional, table-top microcentrifuge and all steps carried out at room temperature.

Sequencing

To confirm that the DNA plasmids were cloned in the correct orientation, the samples were sequenced by *Genewiz Inc.* The samples were diluted to 50 ng/μL by using sterile water. For each reaction, 8 μL of the DNA plasmid at 50 ng/μL provided in a labeled 0.2 mL PCR tube and cap. After verifying the sequences the DNA was amplified via PCR. The 25 μL PCR reaction set up according to table 1.3 and 1.4.

Cleanup Amplified PCR Reaction: MinElute

Amplified DNA was cleaned using MinElute PCR Purification Kit (*QIAGEN*) because this kit isolates only the DNA fragments between 70-4000 nucleotides. 5 volumes of Buffer PB were added to 1 volume of the PCR reaction and mixed. pH was optimized by 3 M sodium acetate (pH 5.0). The samples were applied to the column and washed and then eluted with 10 μL sterile water. Detail information can be found from the manufacturer.

In Vitro Transcription

The MEGA shortscript high yield transcription kit from *Ambion* was used for *in vitro* transcription. Although the incubation time varies for a specific reaction, the general transcription protocol is summarized in Table 1.6.

The moles of template DNA were calculated from the mass, by using the equation below. The average molecular weight of all four nucleotides (330 g/mol) can be used instead of the exact molecular weight of the sequence (assuming that all four nucleotides are in roughly equal proportions).

Table 1.6: Assembly of *in vitro* transcription reaction. Following components were mixed at room temperature and incubated in a water-bath at 37 °C.

Amount	Component
2 µL	T7 ATP Solution (75 mM)
2 µL	T7 GTP Solution (75 mM)
2 µL	T7 CTP Solution (75 mM)
2 µL	T7 UTP Solution (75 mM)
2 µL	T7 10X Reaction Buffer
2 µL	T7 Enzyme Mix
<8 µL	Template DNA
Water (Nuclease-free) to 20 µL final volume.	
~ 4-6 hr incubation at 37 °C. After incubation add 1 µL of DNase and incubate at 37 °C for additional 15 min.	

MW of template = Ave MW per bases x # of base (x 2 for double-stranded)

Moles of template = Mass of template / MW of template

Because each of our variants has 76 bp, the moles in 1 µg of a 76 bp PCR product (double stranded):

MW of template = 330 g/mol x 76 x 2 = 5.016x10⁴ g/mol

Moles of template = (1 µg) / (5.016x 10⁴ g/mol) = 20 pmol of template

The reaction was assembled in an RNase-free microfuge tube (*Eppendorf*) at room temperature. The contents were mixed thoroughly by gently flicking the tube, and then the mixture was incubated at 37 °C. For most applications, a 4 hr incubation was sufficient, but for some reactions the incubation time was extended up to 6 hr. After sufficient incubation, the DNA

template was removed by adding 1 μ L of TURBO DNase to the reaction and continued the incubation at 37 °C for an additional 15 min.

The reaction was terminated by adding 1/2 volume of Gel Loading Buffer II (*Ambion*). The sample was incubated at 80 °C for 3 min and additional 3 min at room temperature. The products were loaded on a 4-15% Tris-HCl denaturing polyacrylamide gel and the reaction was run at 150 volt (1X TBE Buffer was used).

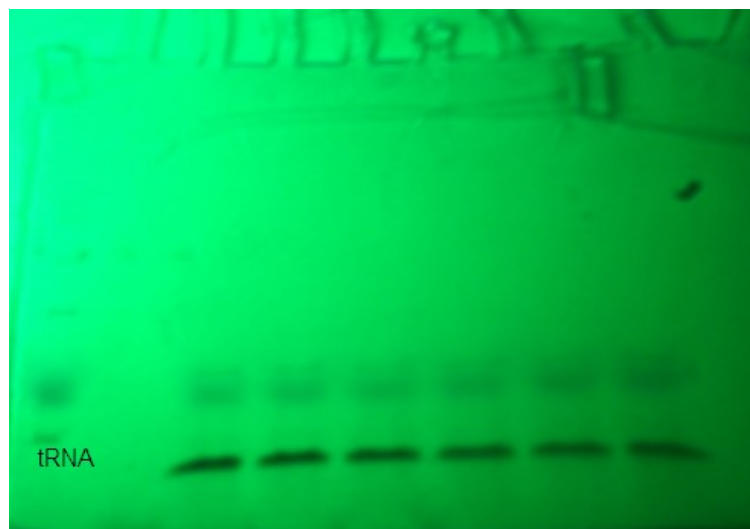


Figure 1.6: Polyacrylamide Gel Electrophoresis (PAGE) of RNA. tRNAs were loaded on a 4-15% Tris-HCl denaturing polyacrylamide gel and the reaction was run at 150 volt for 45 min. The gel was visualized by short wave U.V.

After electrophoresis, the gel was covered with polyvinyl chloride to prevent any contamination. The area of the gel that contains the full-length transcript was visualized using a short wave UV lamp and the area containing the nucleic acid was marked with a pen. The marked-area was cut out with a sterile razor blade and transferred into an RNase-free microfuge tube containing 400 μ L of sterile RNA-free water. The tube was incubated at 4 °C for overnight

with shaking at 10 rpm. Next day, the liquid part was transferred in a new tube and added 6/100 volume of 5 M NaCl and 2.5 volumes ethanol. After vortexing, the tube was incubated at -20°C for at least 6 hr. The sample was centrifuged at 4°C for 45 min at a $21000 \times g$ to pellet the RNA. The supernatant solution was removed carefully and the pellet resuspended in 25 μL of sterile RNA-free water and kept on ice. The tRNA was assessed by UV absorbance using Nanodrop ND-1000A UV-Vis Spectrophotometer.

1.3. RESULTS AND DISCUSSIONS

In vitro techniques were used to investigate sequence structure-function relationships of tRNAs and their interactions with EF-Tu. Binding affinity of EF-Tu to the G:U wobble base distribution on tRNA structure is indirectly inferred based on the ability to participate in protein translation. The T and Acceptor stems of tRNA sequences were selected to mutate at sites predicted to be responsible for weak binding or strong binding.

We introduced G:U mutations to different regions of the tRNA and demonstrated the importance of G:U wobble base-pairs in tRNA participation in translation. We anticipate that our study can be further applied to biomedicine since many new therapeutics require the incorporation of unnatural amino acids into peptides or proteins. In the current study, we used Protein synthesis Using Recombinant Elements (PURE) *in vitro* translation system to demonstrate participation of mutant tRNAs during peptide translation. To identify efficiency of tRNAs in translation, we charged wild type and all mutant tRNAs with Glutamate (a weak amino acid) or Glutamine (a strong amino acid).

Determine G:U mutations in Valine-tRNA

The G:U pair is a fundamental unit of nearly all tRNAs. We introduced individual mutations to specific regions of the tRNAs that interact with EF-Tu. Therefore, we introduced mutations to the acceptor and T stem of tRNA and created G:U wobble base-pairs not naturally present in the Val-tRNA species. Those mutations are highlighted in red in figure 1.7. In addition, to determine the effect of G:U pairing at specific sites, we also introduced a mutation to remove the G:U pair in wild type Val-tRNA at 50:64 position.

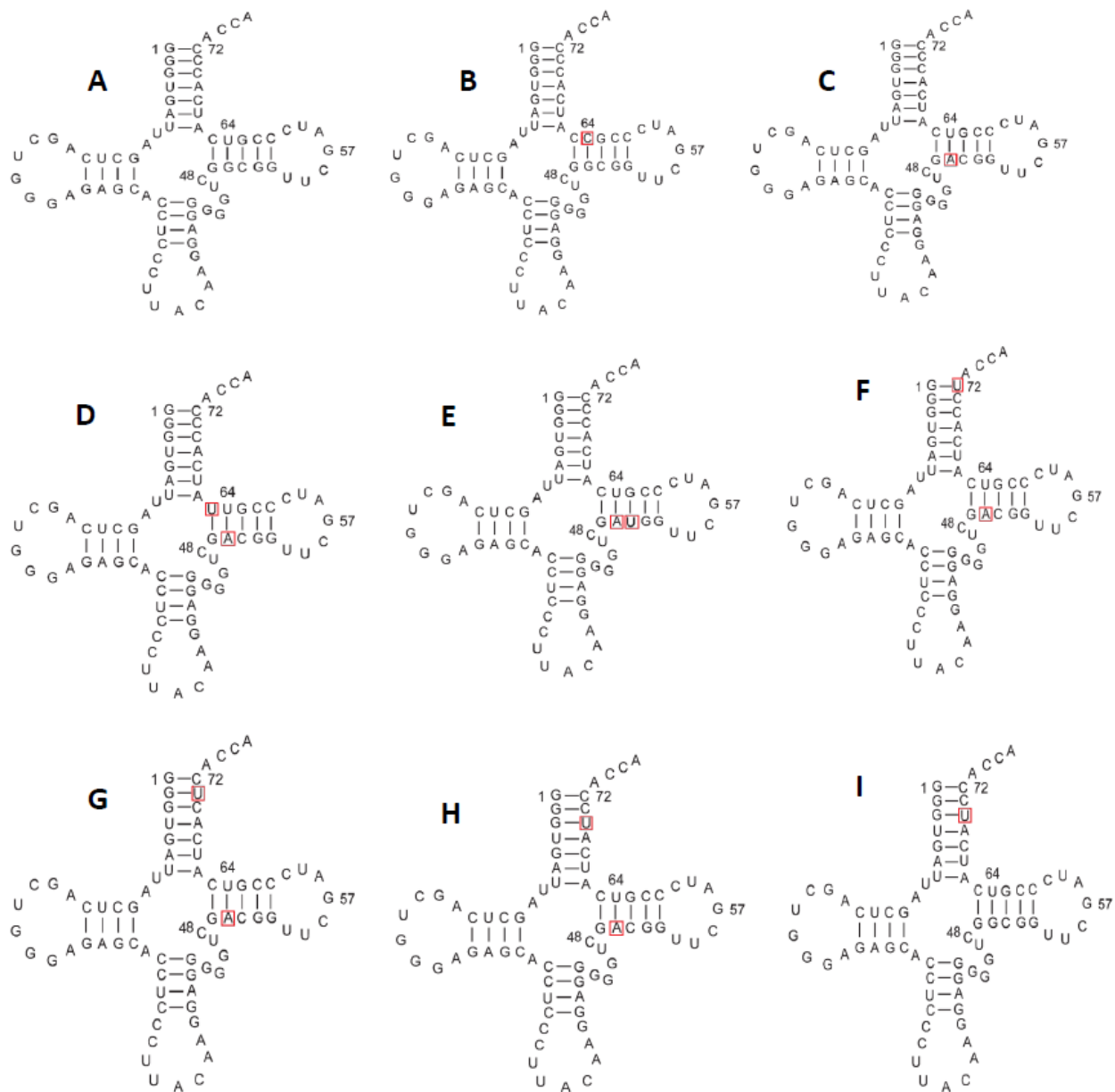


Figure 1.7: Engineered Val-tRNA variants. G:U mutations (highlighted in red) were introduced to the A and T stem of Val-tRNA. Most of mutations were C→U, except some where G→A or U→C.

Cloning and Amplification of DNA

PCR primers were designated as template and amplifying primers. T7 promoter was introduced in order to transcribe DNA into RNA. Designated primers were synthesized by IDT. Although PCR conditions are described in materials and methods, first template primers were amplified for four cycles and then amplifying primers were added to amplify template DNA for an additional 30 cycles. The optimal amplification conditions were determined by gradient PCR and template was amplified using standard PCR conditions which are also described in material and methods section. Optimal annealing temperature displayed diversity, but highest products were mostly captured in 50-52 °C temperature range. The PCR products were cleaned up using a QIAgen gel extraction kit. The length of our DNA was the standard *E. coli* Val-tRNA length which was 76 base pair plus the length of the T7 promoter, 93bp PCR product in length.

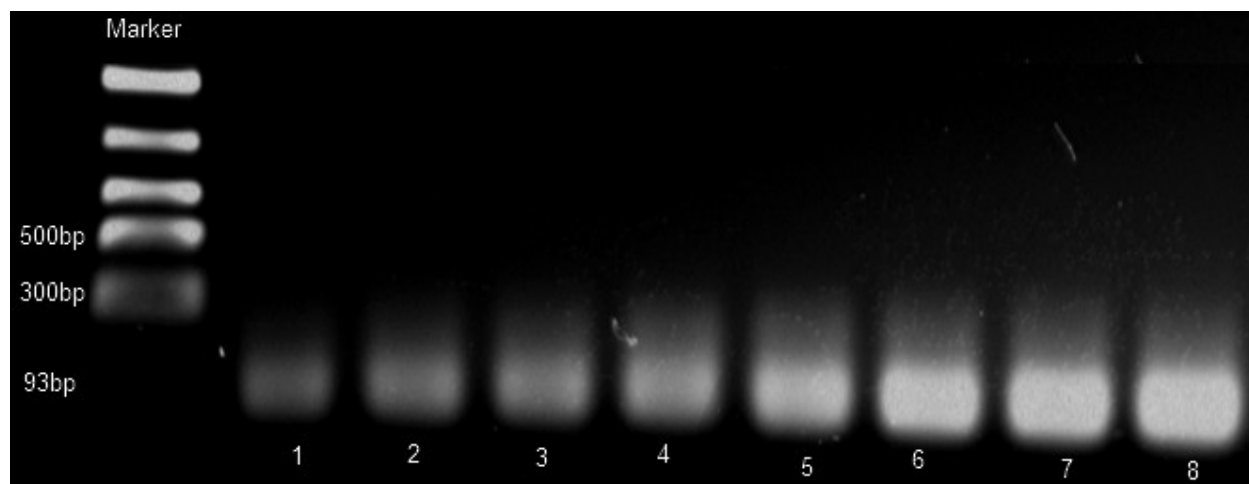


Figure 1.8: Gradient PCR. On the left side marker is shown and the PCR product corresponding to 93bp. The annealing temperature gradient was from lanes 1 to 8 was 60°C, 59.4°C, 58.3°C, 56.3°C, 53.9°C, 52°C, 50.7°C, 50°C, respectively. The higher products were captured at 50-52°C temperature range. Therefore the annealing temperature for PCR amplification of DNA was 52°C.

Purified-PCR product was successfully ligated into the TOPO vector using TOPO TA Cloning Kits (Invitrogen). This kit provided efficient cloning for our direct insertion of amplified PCR products into the plasmid vector. Basically, single T overhangs for TA cloning is possible because the plasmid vector is already linearized. Then Topoisomerase I covalently binds to the vector and Taq polymerase-added single A to the 3' ends of PCR products. This allowed PCR inserts to ligate efficiently with the vector.

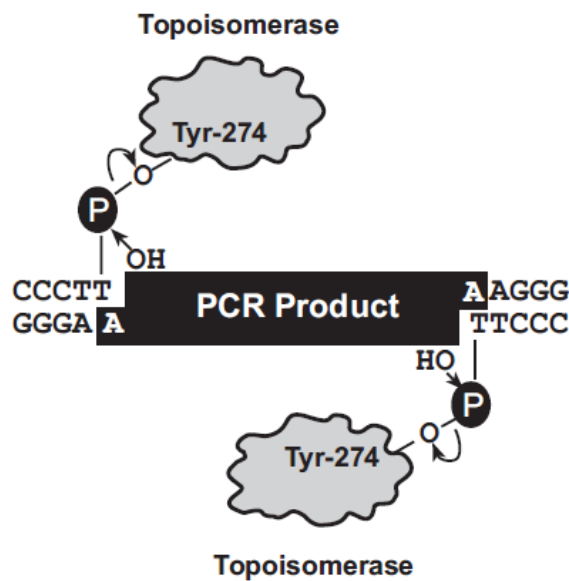


Figure 1.9: Insertion of PCR product into Topo Vector. Topoisomerase-I binds to duplex DNA and cleaves the phosphodiester backbone after 5'-CCCTT in one strand. The phospho-tyrosyl bond between the DNA and enzyme can be attacked by the 5' hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase (TopoTA manual, invitrogen; Shuman 1994).

To analyze constructs, plasmid DNA was purified and DNA sequences were verified by sequencing (*Genewiz Inc*) by using the M13 Forward (-20) and M13 Reverse primers. Verified plasmid DNA was used as a template DNA and amplified by PCR (Figure 1.10). The PCR

conditions were similar to conditions that were summarized in materials and methods in table 1.3 and 1.4. The only exception was 20 ng/ μ L of verified plasmid DNA was used as a template.

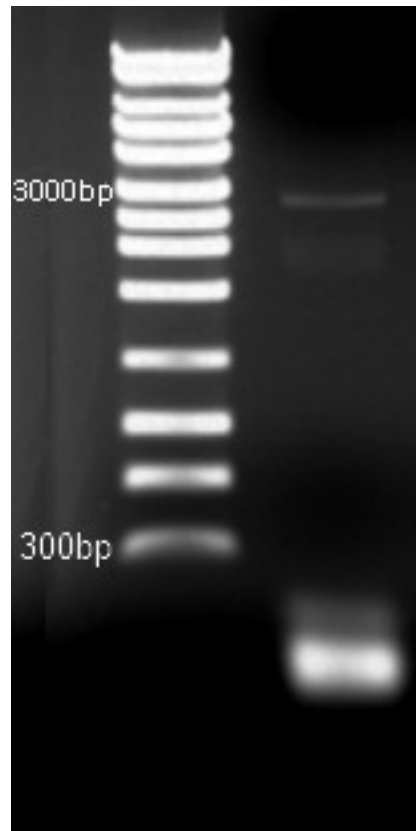


Figure 1.10: Amplified DNA-tRNA. DNA was load on 0.8% Agarose-EtBr gel and visualized by UV. 3000bp correspond template plasmid and the amplified DNA is 93bp.

***In Vitro* transcription**

Wild-type and variants were successfully transcribed into tRNAs. In vitro transcription reaction is described in materials and methods. PCR product that contains a T7 RNA polymerase promoter was used as template for *in vitro* transcription. The template DNA was prepared free of contaminating proteins, nonspecific DNA, and RNA. The components were

assembled at room temperature and incubated at 37 °C for 4-6 hr. In order to get pure RNA product, template DNA was removed by DNase. The transcription reaction was purified by using 4-15% Tris-HCl polyacrylamide gel. To prevent any contamination, each time one polyacrylamide gel was used only for one variant of Val-tRNA. The amount of template and incubation time played a major role on transcription efficiency. Low amount of template DNA caused low yield or no product, using higher amount of template caused having unspecific products (Figure 1.11).

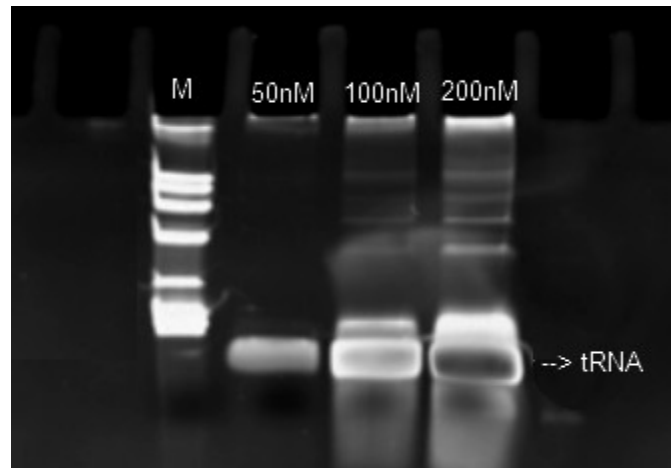


Figure 1.11: *In vitro* transcription: wild-type Val-tRNA by using different amount of template DNA 50 nM, 100 nM, and 200 nM respectively. The Marker (2 ul of Century Size Markers, *Ambion*) and samples were mixed with 10 uL of Gel Loading Buffer II. The samples were incubated at 80 °C for 3 min and 3 additional min at room temperature. The products were loaded on a 4-15% Tris-HCl denaturing polyacrylamide gel and the reaction was run at 150 volt (1X TBE Buffer). After electrophoresis, the gel was incubated in 100 mL water that contained 100 uL diluted EtBr for 5 min and visualized by UV.

After standard mini electrophoresis, the transcription product was visualized by short wave UV and cut out with a sterile razor blade and transferred into sterile RNase-free water. Overnight incubation in water allowed the tRNAs to migrate from the gel slices into the water. The RNA was recovered by precipitation using 100% ethanol and eluted with RNase-free water.

In some cases, low product of tRNA made it difficult to see the pellet, so the supernatant was carefully removed immediately after centrifugation. The tRNA was quantified by using Nanodrop ND-1000A UV-Vis Spectrophotometer (Figure 1.12). Transcription yields varied depending on the variant and the template quality. For low yield reactions, the number of reactions was increased to get adequate product for use during in vitro transcription.

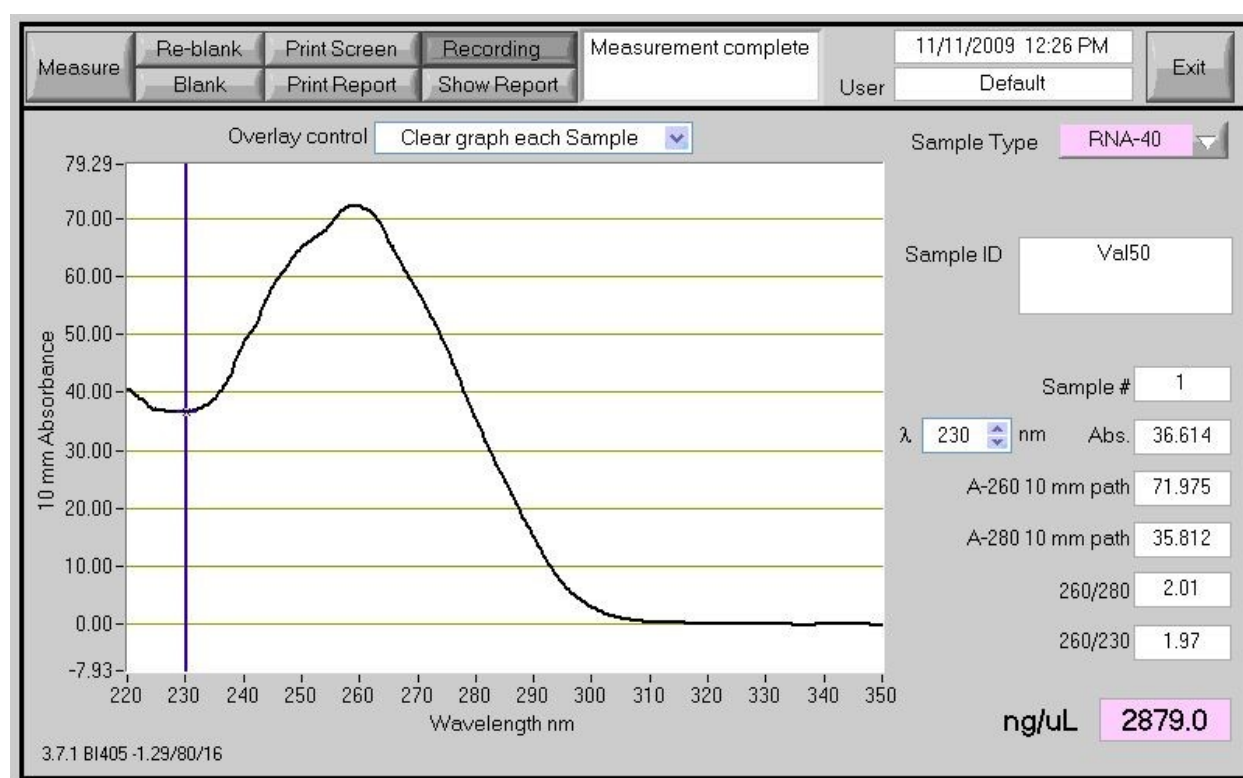


Figure 1.12: Quantification of in vitro transcription product. The tRNA was assessed by UV absorbance using Nanodrop ND-1000A UV-Vis Spectrophotometer.

***In Vitro* Translation (PURE system)**

Affinity of mutant tRNAs to EF-Tu were indirectly assessed by quantification of peptide produced using the PURE system. Except for the ribosomes and tRNAs, the PURE system contains all of the *E. coli* components necessary for protein translation (Shimizu, Y., *et al.*,

2001). *In vitro* translation assay was performed with our collaborator Dr. Matthew Hartman and his laboratory at Virginia Commonwealth University.

Assay with Glutamate (Glutamic Acid)

All tRNAs were charged with glutamate which has been identified as a weak amino acid. To identify efficiency of mutant tRNAs in translation, we assayed the incorporation of each mutant tRNA at a single site in a short peptide containing an N-terminal His-tag. We have used a template MHHHHHHMVHM, where V is the codon that we were interested in covering with one of the mutant tRNAs or wild-type. Wild type Val-tRNA was used as control and the yield of reaction (amount of produced peptide) was determined as 100% and all other results are relative to the wild-type. We tested the entire mutant tRNAs and compared them with wild-type in vitro transcribed Val-tRNA. Figure 1.13 shows results for Val-tRNAs^{Glu} variants. Because glutamate itself is a weak amino acid and Val-tRNA shows moderate or weak affinity to EF-Tu, we did not anticipate a high yield of peptide in comparison to a strong binding amino acid on the weak tRNA. Our result with wild-type Val-tRNA performed as predicted. If we look at the secondary structure of wild-type Val-tRNA, we notice that there is only one G:U base-pair (position 50-64) in the T stem at a position where we predict that G:U would decrease binding affinity of tRNA to EF-Tu. To verify our prediction that a G:U pair at the 50-64 position would decrease the interaction with EF-Tu, we introduced mutations to those sites to remove the G:U pair. Results were parallel with our prediction in which removal of the G:U at those sites would increase the amount of peptide produced. We then moved to position 49-65 on the T stem of the tRNA. We changed C to U at 65 position resulting in a G:U at 49:65. Note, due to the G:U at 50:64, we also mutated pair 50:64 to get rid of the G:U in this variant [The same applies for many of the

variants we produced]. Translation yields nearly doubled for this variant and interestingly this was the highest yield among the entire mutant tRNAs that were mis-acylated with Glutamate. We also introduced a G:U mutation to the 51:63 pair and this resulted in a low amount of peptide produced. Overall, results in the T stem of tRNA showed that G:U lead to a decrease in participation of tRNA during translation at positions 50:64 and 51:63. Conversely, G:U in position 49:65 increased the participation of tRNA during translation. Previous studies support our findings for the T stem of tRNA (Uhlenbeck *et al.*, 2009).

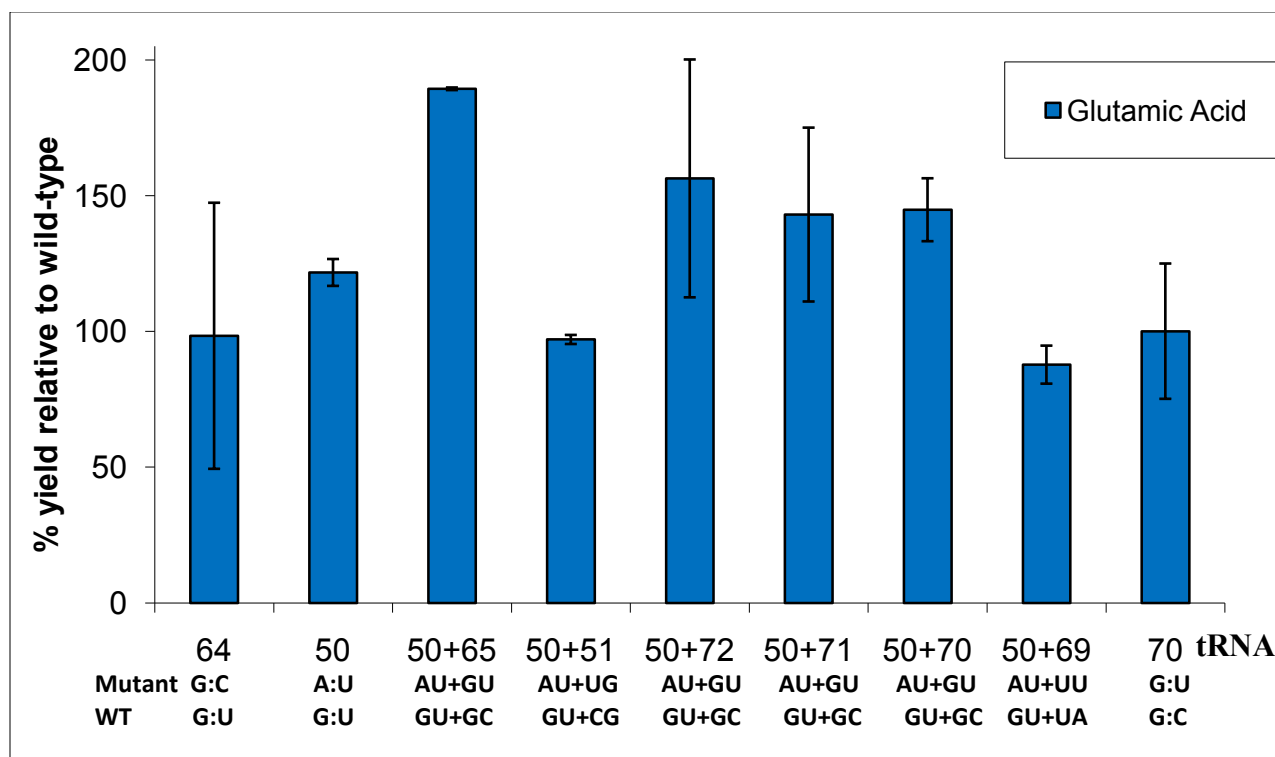


Figure 1.13: Peptide synthesis assay with Val-tRNA variants charged with glutamate. Average yields for successful translations producing peptides. We used an mRNA template corresponding to a MHHHHHHMVHM peptide where V served as our site of interest for incorporation of our mutant tRNAs. We tested wild type and all of the mutant Val-tRNAs, by charging with glutamate using a Flexizyme and the *in vitro* PURE translation system. We measured the yield of the peptides produced. Wild type Valine tRNA was used as control and the yield of reaction (amount of produced peptide) was used as a reference of 100% and all other results relative to the wild type. Numbers on row represent tRNA sites where we introduced mutations.

In addition to the T stem of tRNA, we also introduced G:U mutations to the A stem of tRNA because EF-Tu binds to the A and T stems of tRNA. We introduced mutations at position pairs 1:72, 2:71, and 3:70. In all these three cases, a G:U pairing increased peptide production nearly 1.5 fold in comparison to the wild type. As mentioned above; beside these mutations, we kept the mutation at position 50 which removed the G:U in order to understand the context-dependent affects of these mutations.

Assay with Glutamine

To further dissect the system, we performed the same assay but with glutamine, which is known as a strong-binding amino acid in terms of its interactions with EF-Tu. Therefore, we predicted more tRNA participation in translation in comparison of glutamate. Again, wild type and mutant tRNAs were aminoacylated with glutamine and the same mRNA template was used. As we performed in the assay with glutamate, wild type Val-tRNA was used as control and the amount of produced-peptide was referenced as 100% and all other results are presented relative to the wild type glutamate. We analyzed all of the mutant tRNAs and compared them with the wild-type in vitro transcribed Val-tRNA. Figure 1.14 shows the results for Val-tRNA^{Gln} variants. The assay with glutamine resulted in similar patterns as the assays conducted with glutamate – the only exception being the higher yields due to the fact that glutamine is a strong-binding amino acid. Again, a G:U pairing at position 50:64 and 51:63 appeared to decrease affinity of Val-tRNA^{Gln} to EF-Tu as measured by peptide produced. Conversely, G:U pairings at position 49:65 on the T stem and positions 1:72, 2:71, 3:73 on the acceptor stem increased the ability of the mutant Val-tRNA^{Gln} to participate in translation.

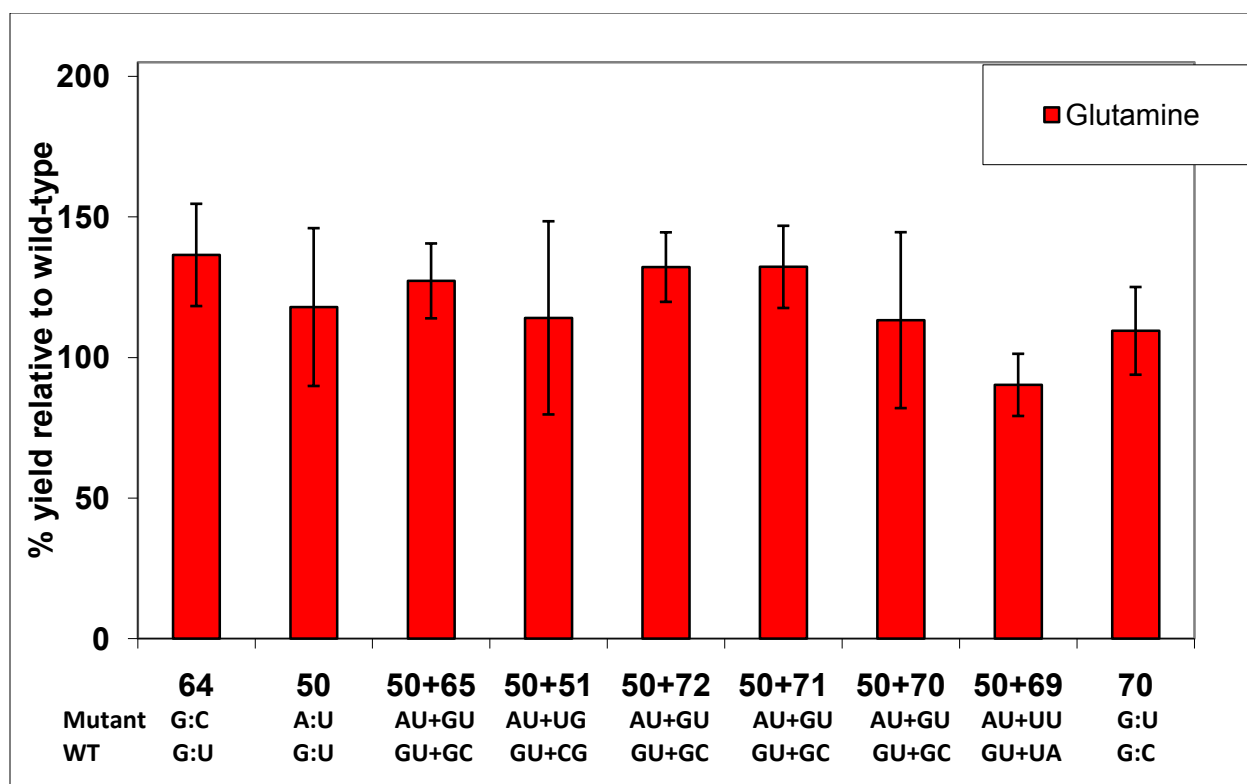


Figure 1.14: Peptide synthesis assay with wild-type and mutant Val-tRNAs charged with glutamine. Average yields for successful translations producing peptides. We used an mRNA template corresponding to a MHHHHHMHVM peptide where V served as our site of interest for incorporation of our mutant tRNAs. We tested wild type and all of the mutant Val-tRNAs, by charging with glutamine using a Flexizyme and the *in vitro* PURE translation system. We measured the yield of the peptides produced. Wild type Valine tRNA was used as control and the yield of reaction (amount of produced peptide) was used as a reference of 100% and all other results relative to the wild type. Numbers on row represent tRNA sites where we introduced mutations.

1.4. CONCLUSION

Based on overall results, we have demonstrated that the G:U wobble pair plays an important role in participation of aminocylated-tRNA in the translation system. We have mapped a G:U distribution onto the tRNA structure based on our analysis (Figure 1.15). For the most part, a G:U wobble base-pair in the acceptor stem increases participation of aminoacylated-tRNA in protein synthesis; whereas, a G:U pair in the T stem decreases participation with the exception at position 49:65 which enhances participation.

Although our results are mostly consistent with Uhlenbeck's thermodynamic compensation model, there are some minor variations. These variations might be due to the fact that we studied a different species of tRNA. In previous studies, Uhlenbecks and colleagues did not specify which Val-tRNA species they used thus making a direct comparison difficult. We will need to conduct further investigations with different tRNA species in order to establish a hypothesis based on effects of the G:U wobble base-pair in tRNA participation during translation.

We are confident that our results can next be exploited to generate tRNA species charged with unnatural amino acids with the intention that the engineered species will enhance incorporation of the unnatural amino acid during peptide synthesis. Such experiments are currently being set-up in the Gaucher group.

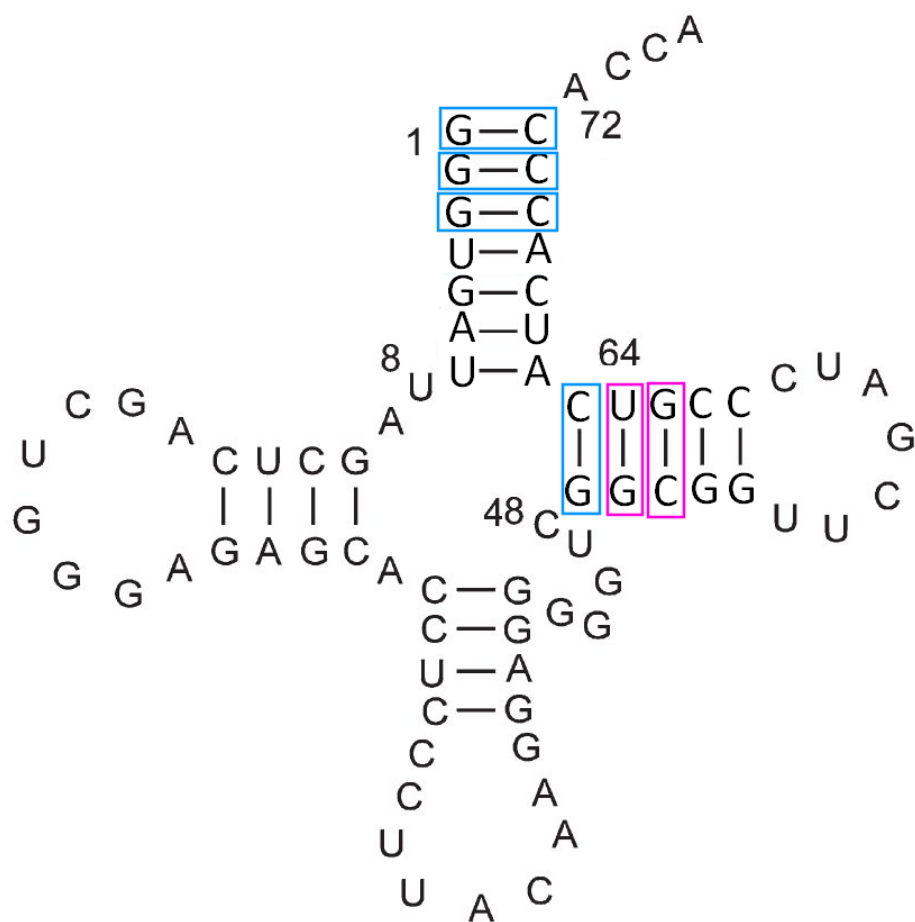


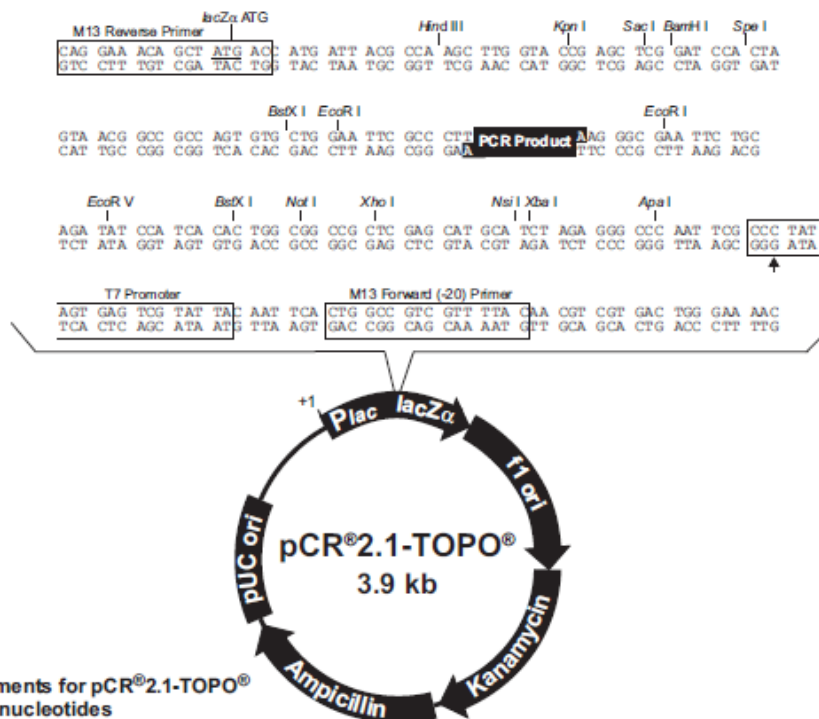
Figure 1.15: G:U base pair distribution on tRNA based on participation in peptide translation. The representative tRNA is the wild-type Val-tRNA. The regions that appear to increase participation are highlighted in cyan. Pink sites represent regions where the G:U decreases participation.

1.5. APPENDIX

Map of pCR[®]2.1-TOPO[®]

pCR[®]2.1-TOPO[®] Map

The map below shows the features of pCR[®]2.1-TOPO[®] and the sequence surrounding the TOPO[®] Cloning site. Restriction sites are labeled to indicate the actual cleavage site. The arrow indicates the start of transcription for T7 polymerase. The complete sequence of pCR[®]2.1-TOPO[®] is available for downloading from our Web site (www.invitrogen.com) or by contacting Technical Service (page 24).



Comments for pCR[®]2.1-TOPO[®] 3931 nucleotides

LacZα fragment: bases 1-547
 M13 reverse priming site: bases 205-221
 Multiple cloning site: bases 234-357
 T7 promoter/priming site: bases 364-383
 M13 Forward (-20) priming site: bases 391-406
 f1 origin: bases 548-985
 Kanamycin resistance ORF: bases 1319-2113
 Ampicillin resistance ORF: bases 2131-2991
 pUC origin: bases 3136-3809

1.6. BIBLIOGRAPHY

- Crick, F. H., L. Barnett, et al. (1961). "General nature of the genetic code for proteins." *Nature* **192**: 1227-1232.
- Crick, F. H. (1958). "On protein synthesis." *Symp Soc Exp Biol* **12**: 138-163.
- Dale, T. and O. C. Uhlenbeck (2005). "Amino acid specificity in translation." *Trends Biochem Sci* **30**(12): 659-665.
- Dale, T., L. E. Sanderson, et al. (2004). "The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid." *Biochemistry* **43**(20): 6159-6166.
- Gabriel, K., J. Schneider, et al. (1996). "Functional evidence for indirect recognition of G.U in tRNA(Ala) by alanyl-tRNA synthetase." *Science* **271**(5246): 195-197.
- Hobom, B. (1980). "Surgery of Genes - at the Doorstep of Synthetic Biology." *Medizinische Klinik* **75**(24): 14-21.
- Ibba, M., A. W. Curnow, et al. (1997). "Aminoacyl-tRNA synthesis: divergent routes to a common goal." *Trends Biochem Sci* **22**(2): 39-42.
- Schrader, J. M., S. J. Chapman, et al. (2009). "Understanding the sequence specificity of tRNA binding to elongation factor Tu using tRNA mutagenesis." *J Mol Biol* **386**(5): 1255-1264.
- Ladner, J. E., A. Jack, et al. (1975). "Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution." *Proc Natl Acad Sci U S A* **72**(11): 4414-4418.
- LaRiviere, F. J., A. D. Wolfson, et al. (2001). "Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation." *Science* **294**(5540): 165-168.
- RajBhandary, U. L. (1994). "Initiator transfer RNAs." *J Bacteriol* **176**(3): 547-552.
- Rawls, R. L. (2000). "'Synthetic biology' makes its debut." *Chemical & Engineering News* **78**(17): 49-+.
- Shimizu, Y., A. Inoue, et al. (2001). "Cell-free translation reconstituted with purified components." *Nature Biotechnology* **19**(8): 751-755.
- Sprinzel, M. (2006). "Chemistry of aminoacylation and peptide bond formation on the 3' terminus of tRNA." *Journal of Biosciences* **31**(4): 489-496.
- Szybalski, Waclaw. In Vivo and in Vitro Initiation of Transcription, Page 405. In: A. Kohn and A. Shatkay (Eds.), Control of Gene Expression, pp. 23-4, and Discussion pp. 404-5 (Szybalski's concept of Synthetic Biology), 411-2, 415-7. New York: Plenum Press, 1974

Woese, C. R. (2001). "Translation: in retrospect and prospect." *RNA* 7(8): 1055-1067.

Xu, D., T. Landon, *et al.* (2007). "The electrostatic characteristics of G center dot U wobble base pairs." *Nucleic Acids Research* 35(11): 3836-3847.

Varani, G. and W. H. McClain (2000). "The G center dot U wobble base pair - A fundamental building block of RNA structure crucial to RNA function in diverse biological systems." *Embo Reports* 1(1): 18-23.

Yarus M, Smith D. 1995. In tRNA: Structure, Biosynthesis and Function, ed.DSoll, UL RajBhandary, pp. 443–69. Washington, DC: *ASM Press*.

CHAPTER-II

2. FUNCTIONAL DIVERGENCE AMONG HOMOLOGOUS PROTEINS

2.1. BACKGROUND AND SIGNIFICANCE

Functional Divergence

Members of functionally-diverse homologous protein families can be used to study structural determinants of protein function and the relationship between evolving protein structure and function. Functional diversification within a protein family should most often correspond to changes in protein sequence. Functionally important amino acid residues are conserved throughout evolutionary history. Alternatively, a change in amino acid conservation at a specific site may be a sign of functional divergence. Most research groups in the molecular evolution community focus their efforts on sites in a sequence and attempt to understand how historical mutations drive functional diversification. This focus is a consequence of the truism that the selective constraint acting at a site is inversely related to evolutionary rate of change at that site (Gaucher *et al.*, 2002). Following this idea, statistical models have been developed by many groups to understand and test functional divergence within a protein family (Gu, 2003-2006; Wang *et al.*, 2006; Knudsen *et al.*, 2001).

Rate Heterogeneity and Heterotachy

Rate heterogeneity is a well-established pattern used by evolutionary biologists to dissect a collection of sequences. Rate heterogeneity is defined as unequal mutation rates among sequence positions within a collection of aligned sequences. The substitution rate of sites in a given gene family can change through their evolutionary history. Therefore, this variability in evolutionary rate can have an impact on analyses such as those that infer phylogenetic tree

generation, which is important for the study of evolution (Revell *et al.*, 2008). Rate heterogeneity across sites is captured by a so-called gamma distribution and its shape parameters, α ; basically, an increase in α corresponds to a decrease in the rate variation across sites (Goldman and Whelan, 2000). For example, a set of four homologous genes of four amino acids positions from four different species is shown in figure 2.1. If we assume that each square represents one unit of evolutionary change, the site-specific substitution rate will be different due to unequal mutation rates among the sequence positions. For instance, in the representative scheme, the first two positions represent a low evolutionary rate and they are constant throughout the history of the four species. In the third position, the mutation rate is moderate. The fourth position, however, represents a rapidly evolving site and may be indicative of altered selective constraints acting at this position.

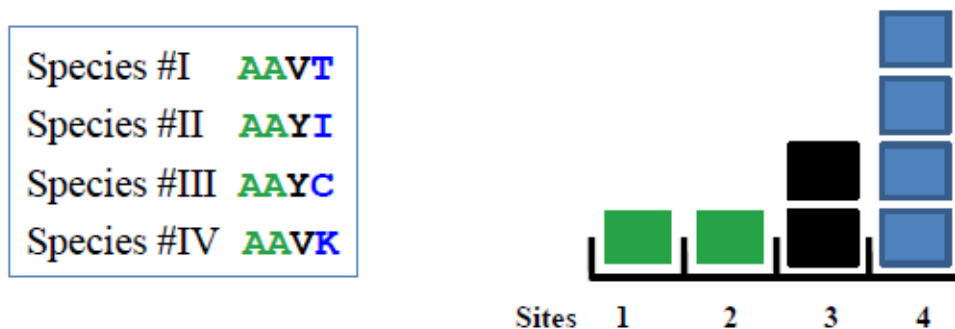


Figure 2.1. A hypothetical scheme showing rate heterogeneity of sites within homologous genes from four species. Left, hypothetical amino acid alignment for the given gene. The first two amino acid positions are identical but the last two are variable. Right, each square represents one unit of evolutionary change, the first two positions represents low evolutionary rates. The third position represents a moderate amount of mutation, or evolutionary change. The fourth position represents a rapidly evolving site.

Previous studies have shown that variable rates at specific sites can cause changes in the structure and function of a protein (Miyamoto and Fitch 1995; Philippe and Lopez 2001). The traditional view of protein sequence change was that the replacement rate at individual sites stays

constant throughout evolutionary history, which is called homotachous model. This model presumes that conserved sites are always evolving slowly, and that non-conserved sites are rapidly evolving across the entire phylogenetic tree (Yang 1996a).

For example, as shown in the figure 2.2., the first and third sites are evolving slowly, the second site is moderately evolving, and the fourth is rapidly evolving. Under the homotachous model, we typically assume that this process remains constant throughout evolutionary history. Therefore, the last common ancestor of life evolved into bacteria and eukaryotes and the mutation rate stayed the same throughout time, so the first site remained slow-evolving in both bacteria and eukaryotes and the last site remained fast-evolving in bacteria and eukaryotes as well.

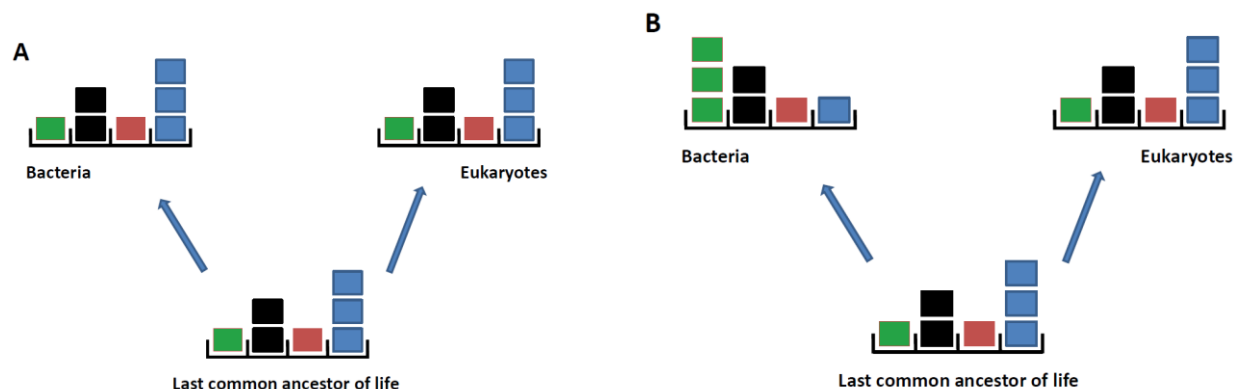


Figure 2.2. Homotachous (A) and heterotachous (B) rate heterogeneity models. The last common ancestor of life diverged to give rise to bacteria and eukaryotes. Homotachy can be considered a form of purifying selection whereas Heterotachy can be considered a form of diversifying selection.

However, later studies have shown that a homotachous model is insufficient to capture the site specific rate shift for protein-coding sequences (Fitch and Markowitz, 1970; Miyamoto and Fitch, 1995; Gaucher *et al.*, 2001). These studies implied that selection at the organismal

level can be variable at particular sites of DNA and protein. Specifically, this means that positions have variable rates of mutations in different branches of an evolutionary tree (Gaucher *et al.*, 2001). Again, in the representative figure, we have the same common ancestor and distribution of site-specific rates (two slow, one moderate, and one rapid). However, this time, after the split that gives rise to bacteria and eukaryotes, the first site changes from slowly evolving to now rapidly evolving, and the fourth site goes from rapidly evolving to now slowly evolving. Thus, these shifts in rate may have arisen in response to a shift of functional constraints operating on this gene as it diverges between bacteria and eukaryotes and may lead to minor modifications in its functionality or properties between these two domains of life.

Evolutionary biologists have developed site-specific rate shift models under both the heterotachous and homotachous model framework to determine which model is the most suitable for a given dataset (Philippe, H. *et al.* (2000); Gaucher *et al.* (2002); Cole *et al.* (2010); Kolaczkowski *et al.* (2008)). Gaucher *et al.* have compared Elongation Factor (EF) gene families of bacteria and eukaryotes under the non-homogeneous rate heterogeneity model (Figure 2.3). They analyzed EF sequences using the DIVERGE software (Gu *et al.*, 2002) to identify evolutionary rates of sites which have undergone site-specific rate shifts. They have identified some sites that rapidly evolve in bacteria but slowly evolve in eukaryotes and vice versa. They showed the distribution of sites associated with a shift in evolutionary rates between bacteria and eukaryotes. For example in the figure 2.3., green shows the sites that are rapidly evolving in bacteria but slowly evolving in eukaryotes. Conversely, the sites highlighted in red are rapidly evolving in eukaryotes but slowly evolving in bacteria. The goal then is to make functional inferences, as well as, propose functional explanations about how these patterns arose throughout evolutionary history.

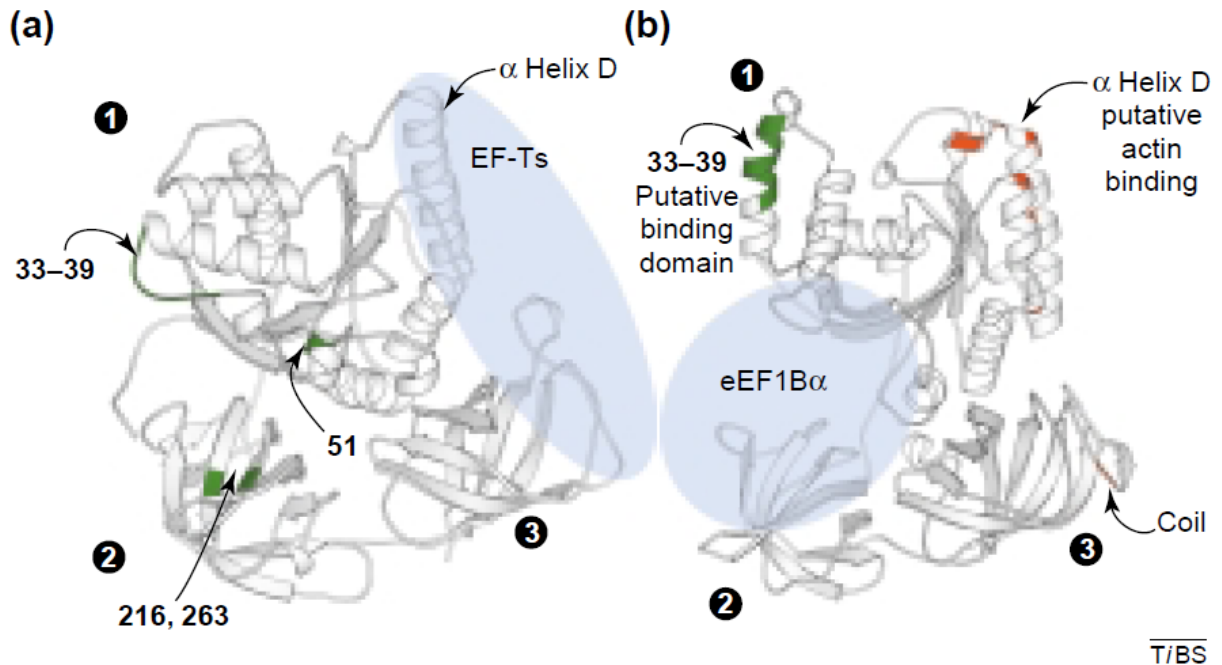


Figure 2.3. Distribution of heterotachous sites onto the three dimensional structure of EF-Tu (a) and eEF1A (b). Nucleotide exchange factors (EF-Ts and eEF1B) are represented in circle. Green identifies sites that are evolving more slowly in eukaryotes relative to bacteria, and red sites are rapidly evolving in eukaryotes than in bacteria. Different domains of the proteins are numbered 1-3 (Gaucher *et al.* (2002)).

One form of rate heterogeneity is known as heterotachy - which is an important process of protein evolution and a signature of functional divergence. Heterotachy comes from Greek language and refers to differences in speed (of mutations). The heterotachy model allows the mutation rate at a position to vary in different branches of the evolutionary tree and this phenomenon is routinely found among homologous sequences of distantly related organisms, where the homologous proteins have varying functions (Lopez *et al.*, 2002). Statistical and empirical analyses have shown that variation within sites can influence phylogenetic tree making

(Lopez, *et al.*, 1999; Philippe, *et al.*, 2000; Gaucher *et al.*, 2002). Some residues might be subject to altered functional constraints in various portions of a phylogenetic tree due to changes in the function of the protein and these changes in evolutionary conservation can misguide phylogenetic tree building algorithms. Figure 2.4 shows how this pattern is manifested at the protein sequence level. For example, a particular site is occupied by lysine and is thus slowly evolving or highly conserved within eukaryote eEF1A proteins. This same position within bacterial EF-Tu, however, is occupied by numerous amino acids. We therefore may conclude that there is a strong selective constraint acting on this site in eukaryotes but this constraint has been eliminated or diminished in bacteria.

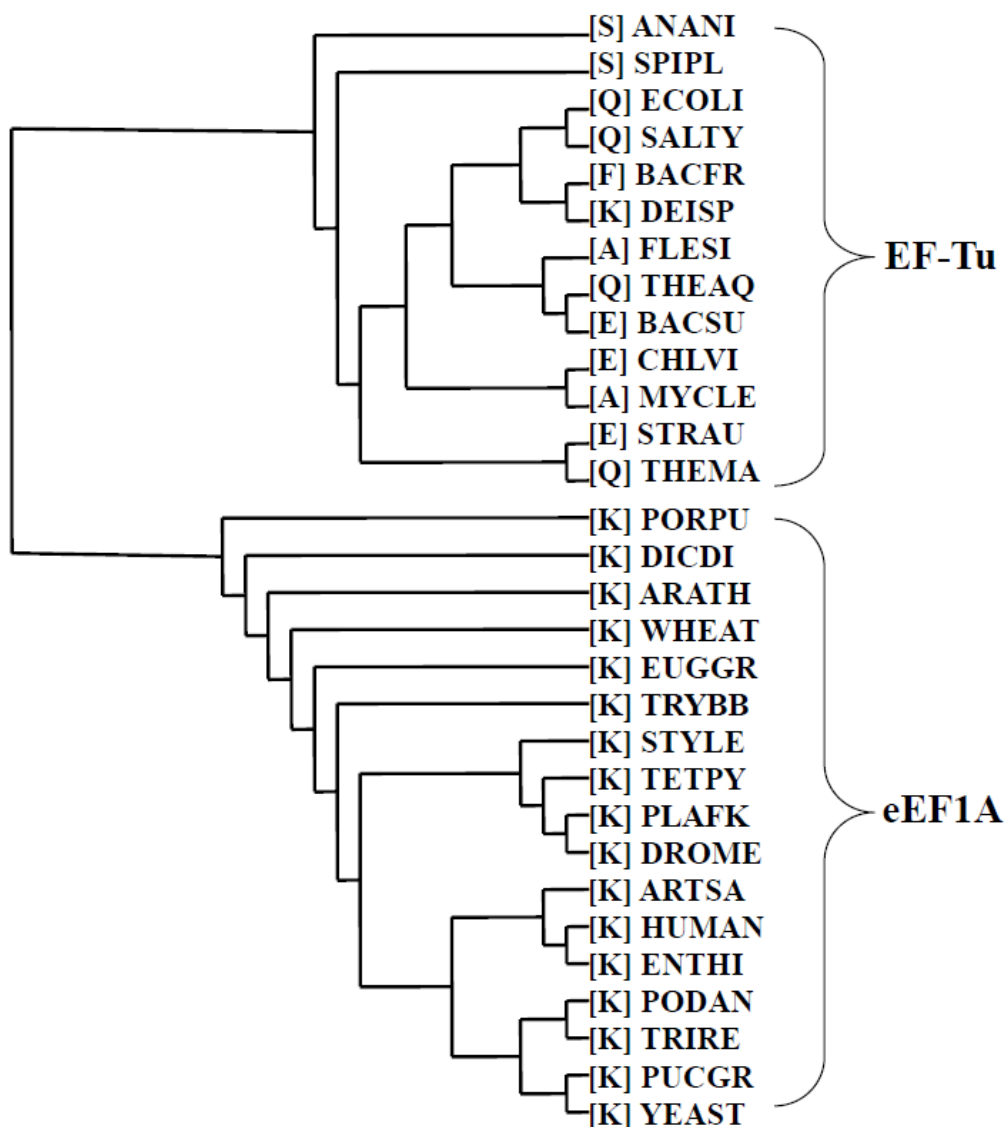


Figure 2.4. An example of heterotachy at a particular site (aligned-position 153) in Elongation Factor gene family. Lysine is present universally in Eukaryotes whereas numerous amino acids occupy this site in Bacteria.

There are three types (type-0, type-I, type-II) of functional divergence that are associated with site-specific rate shifts (Gu, 2001). These types are based on the assumption that sites which are important for function of a protein tend to be conserved throughout evolutionary history (Cole *et al.*, 2010). Type-0 represents conserved residues that are critical in function and are

shared by all members of a gene family. Type-I functional divergence represents amino acid configurations that are conserved in one portion of a phylogeny but variable in another portion of the phylogeny (also referred to as Heterotachy). Type-II functional divergence represents amino acid configurations that are conserved among all portions of a phylogeny but in which the exact amino acid is different among the portions of the tree (for example, positively charged residue versus negatively charged residue)

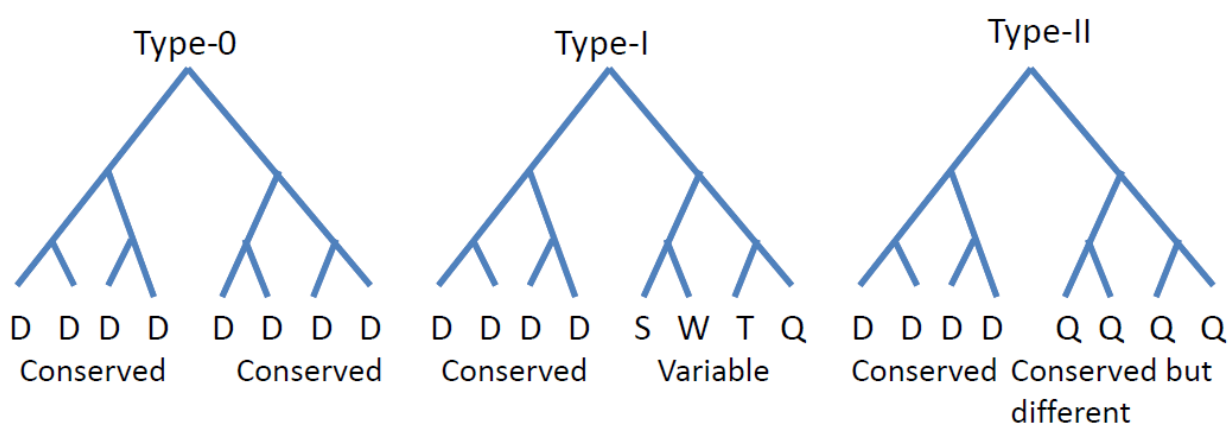


Figure 2.5. Diagrams of site-specific patterns of functional constraints; type-0, type-I and type-II amino acid configurations. Type-0 functional divergence is universally conserved throughout the whole gene family. Type-I functional divergence in which a specific site is occupied by a conserved aspartic acid (D) residue in one lineage of the tree but another lineage is occupied by many different types of amino acids residues. Type-II functional divergence in which one lineage of homologous proteins is occupied by a conserved aspartic acid residue (D) while the other lineage is also occupied by a conserved (but different) amino acid residue (glutamine, Q). (Modified from Cole and Gaucher, 2010)

Predicting Functional Divergence

In the present study, we analyzed such patterns of functional divergence at the protein level. We compared Elongation Factor (EF) proteins between the bacterial and eukaryotic domains of life. We identified type-I and type-II functionally divergent sites in the EF gene family and generated mutant variants in attempt to experimentally validate computational

predictions of heterotachy. To date, no such study has been published, thus we are under the assumption that this is a novel research direction and will provide biological explanations for patterns of sequence conservation/variability.

Elongation Factor Tu

Elongation Factor Tu (EF-Tu) is a bacterial GTP-binding protein composed of three structural domains (Kjeldgaard *et al.*, 1996). EF-Tu is essential for translation in bacteria and its primary role is to bind and deliver the aminoacylated-tRNA (aa-tRNA) complex to the A site of the ribosome during the elongation step of protein synthesis (Lodish, *et al.*, 2008). Binding of the aa-tRNA complex to the ribosome, and subsequent disassociation, are fundamental steps for both efficient and accurate gene expression. This process is catalyzed by the formation of a EF-Tu:GTP:aa-tRNA ternary complex. After binding to the A site of ribosome, GTP is hydrolyzed to GDP which leads to a conformational change (inactive form) in EF-Tu and results in the release of the EF molecule. In order to bind aa-tRNA, EF-Tu has to bind a GTP molecule and this enables the protein to adopt an active confirmation to accept the charged tRNA. To cycle between the active and inactive forms, EF-Tu requires a nucleotide exchange factor nucleotide. In bacteria, this exchange factor is called Elongation Factor Ts (EF-Ts) and it accelerates the rate of GDP release from a EF-Tu molecule in the inactive site. The unoccupied nucleotide binding site in the inactive EF-Tu is then available for GTP binding and this allows the protein to readopt the active confirmation. The nucleotide exchange mechanism itself begins with the binding of EF-Ts to EF-Tu:GDP complex and the reaction results in the ternary complex of EF-Tu:GDP:EF-Ts complex. Due to the fact that this complex is unstable, GDP rapidly dissociates from the complex and GTP binds to the EF-Tu:EF-Ts complex due to the higher affinity of GTP

over GDP within the nucleotide binding site. The nucleotide exchange reaction is completed with the dissociation of EF-Ts from EF-Tu:GTP:EF-Ts and the final complex remains as EF-Tu:GTP. EF-Ts interacts with EF-Tu at the surface of domain-1 and domain-3. The N terminal domain of EF-Ts interacts with domain-1 of EF-Tu, whereas the C terminal of EF-Ts interacts with domain-3 of EF-Tu.

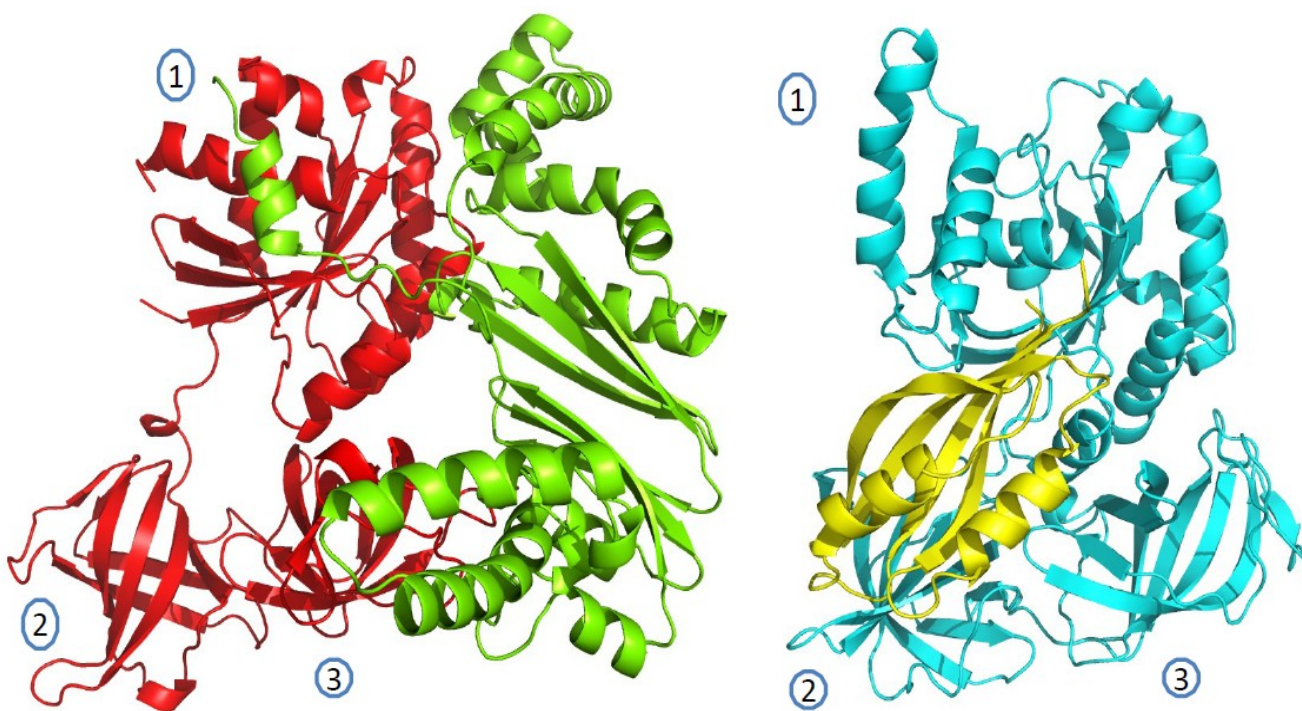


Figure 2.6. Structural organization of EF-Tu (bacteria) and eEF1A (eukaryotes). A) The structure of EF-Tu (Red) from *E.coli* binds to its nucleotide exchange factor EF-Ts (Green). EF-Tu has three domains and EF-Ts binds in domain 1 and 3. B) The structure of eEF1A(Cyan) from *S. cerevisiae* bound to eEF1B (Yellow). eEF1A has three well defined domains just as EF-Tu (indicative of their homologous relationship). Domain I binds GTP, domain II is proposed to bind the aminoacyl end of the aa-tRNA, and domains II and III are linked to actin binding. The nucleotide exchange factor in eukaryotes called eEF1B binds eEF1A in domains I and II. The numbers 1–3 represent the different domains of the proteins. The figures were prepared with *Pymol* using Protein Data Bank (PDB) 1F60 (The Yeast Elongation Factor Complex EEF1A:eEF1B) and 1EFU (Elongation Factor Complex EF-Tu/EF-Ts from *E.coli*).

Elongation Factor 1A

Eukaryotic elongation factor 1A (eEF1A) is a eukaryotic GTP-binding protein and homolog of the bacterial protein EF-Tu (Figure 2.7.). eEF1A thus has a major role in protein synthesis in eukaryotes. The overall properties of eEF1A are analogous to EF-Tu since they are homologous proteins. Some features, however, are different. eEF1A interacts with actin filaments and is thought to use actin as a track system as it shuttles tRNAs to and from the nucleus and a ribosome. Like EF-Tu, eEF1A has to be reactivated with GTP before binding another aminoacylated tRNA. The activation reaction, however, is catalyzed by binding the nucleotide exchange factor eEF1B which is **not** homologous with EF-Ts from bacteria. Recent studies have shown that the nucleotide exchange mechanism of eEF1A is fundamentally different from that of EF-Tu (Andersen *et al.*, 2001). eEF1B interacts with domains 1 and 2 of eEF1A by disrupting the switch 2 region of eEF1A (Figure 2.6) and inserting a lysine in the Mg²⁺ binding site of the GDP/GTP binding site in domain 1 of eEF1A. Conversely, in bacteria, EF-Ts interacts with EF-Tu at the surface of domain-1 and domain-3.

Length=458

Score = 154 bits (389), Expect = 5e-42, Method: Compositional matrix adjust.

Identities = 133/441 (30%), Similarities = 199/441 (45%), Gaps = 60/441 (14%)

Query	8	RTKPHVNVGTIGHVDHGKTTLTAAI-----TTVLAKTYGGAARAFDG-----IDN	52
		+ K H+NV IGHVD GK+T T + + K AA G +D	
Sbjct	3	KEKSHINVVVIGHVDSGKSTTTGHLIYKCGGIDKRTIEKFEKEAAELGKGSFKYAWVLDK	62
Query	53	APEEKARGITINTSHVEYDTPTRHYAVVDCPGHADYVKNMITGAAQMDGAILVVAATDGP	112
		E+ RGITI+ + +++TP V+D PGH D++KNMITG +Q D AIL++A G	
Sbjct	63	LKAERERGITIDIALWKFETPKYQVTVIDAPGHRDFIKNMITGTSQADCAILIIAGGVGE	122
Query	113	MP-----QTREHILLGRQVGVPYIIIVFLNKCDMVD-DEELLELVEMEVRELLSQYDFP	164
		QTREH LL +GV +IV +NK D V DE + + E + + +	
Sbjct	123	FEAGISKDGTREHALLAFTLGLVRQLIVAVNKMDSVKWDESRFQEIVKETSNIKKVGYN	182

Query	165	GDDTPIV-----RGSALKALEGDAEW-----EAKILELAGFLDSY--IPEPERAIDK	209
		P V G + +A W +A +++ L++ I +P R DK	
Sbjct	183	PKTVPFVPISGWNGDNMIEATTNAPWYKGWEKETKAGVVKGKTLLEAIDAIEQPSRPTDK	242
Query	210	PFLLPIDVFSISGRGTVVTGRVERGIIKVGEVEI--VGIKETQKSTCTGVEMFRKLLD	267
		P LP++DV+ I G GTV GRVE G+IK G V G+ KS VEM + L+	
Sbjct	243	PLRLPLQDVYKIGGIGTVPVGRVETGVIKPGMVVTFAPAGVTTEVKS----VEMHHEQLE	298
Query	268	EGRAGENVGVLRLRGIKREEIERGQVL--AKPGTIKPHTKFESEVYILSK--DEGGRHTPF	323
		+G G+NVG ++ + +EI RG V AK K F + V +L+ ++P	
Sbjct	299	QGVPGDNVGFNVKNVSVKEIRRGNVCGDAKNPPKGCASFNATVIVLNLHPGQISAGYSPV	358
Query	324	FKGYRPQFYFR-----TTDVTGTIELPEGVEMVMPGDNIKMVVTLIHPIAMDDGL----	373
		+ R D +L + + + GD + P+ ++	
Sbjct	359	LDCHTAHIACRFDELLEKNDRRSGKKLEDHPKFLKSGDAALVKFVPSKPMCVEAFSEYPP	418
Query	374	--RFAIREGGRTVGAGVVAKV	392
		RFA+R+ +TV GV+ V	
Sbjct	419	LGRFAVRDMRQTVAVGVIKSV	439

Figure 2.7. Comparison of sequence similarity between EF-Tu from *E. coli* and eEF1A from *S. cerevisiae*. Max score 154, total score 154, query coverage 97%, E value 5e-42. Query (EF-Tu), Sbjct (eEF1A).

In addition to an important role in protein synthesis, studies have shown that eEF1A has several other activities, including interactions with the actin cytoskeleton (as mentioned above) (Gross *et al.*, 2005), lipotoxic cell death (Borradaile *et al.*, 2006), ubiquitin-dependent degradation of N-acetylated proteins (Gonen *et al.*, 1994), and quality control of newly synthesized proteins (Hotokezaka *et al.*, 2002).

Although the structures of the eEF1A and EF-Tu are slightly different, they share the same mechanism of disruption within the Mg^{2+} coordination center of the nucleotide exchange factors eEF1B and EF-Ts, again, which are not homologous proteins. Despite intense efforts and large interest in understanding these patterns, we are the first to experimentally validate concepts that attempt to directly connect heterotachy to functional divergence among homologous proteins. Our research attempts to mutate specific sites in order to manipulate the abilities of these elongation factors (Tu and 1A) to bind wild-type nucleotide exchange factors from either

their cognate domain of life or the exchange factor from the other domain of life. We have used EF sequences from thirteen bacterial species and seventeen eukaryotic species (Figure 2.8.). To analyze these sequences, we used the DIVERGE (Detecting Variability in Evolutionary Rates among GENes) program which identifies site-specific rate shifts using a multiple sequence alignment of amino acids in combination with a phylogenetic tree (Gu and Kent, 2001). Site-specific statistical analysis was based on their posterior probability (PP) to determine critical amino acids residues for type-I and type-II functional divergence. We also applied cut-off values in attempt to better understand how the posterior probability associated with whether a site experiencing either Type-I or Type-II divergence may be tied to functional divergence.

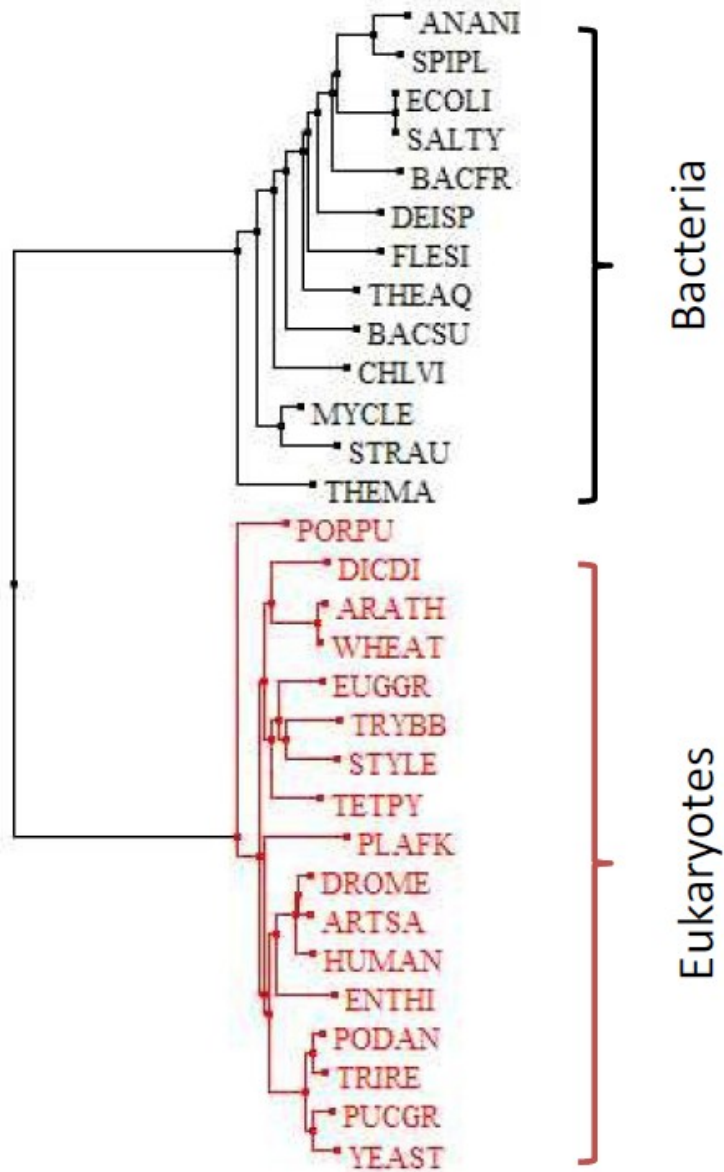


Figure 2.8. The tree used as input for DIVERGE from which the bacterial and eukaryotic clusters were selected. EF-Tu/eEF1A protein phylogeny roots the universal tree between bacteria and eukaryotes. The branches are based on the 394/458 shared amino acid positions (Baldauf *et al.*, 1996).

```
((((((((((ANANI:0.055559,SPIPL:0.042250):0.060464,(ECOLI:0.000000,SALTY:0.000000):0.096684):0.008197,BACFR:0.110337):0.024635,DEISP:0.102696):0.016859,FLESI:0.121341):0.009352,THEAQ:0.088857):0.025117,BACSU:0.114748):0.023583,CHLVI:0.123874):0.025110,(MYCLE:0.034067,STRAU:0.094238):0.037711):0.032692,THEMA:0.123007):0.372402,(PORPU:0.080024,(((DICDI:0.093946,(ARATH:0.013684,WHEAT:0.006198):0.076759):0.012477,((EUGGR:0.075150,(TRYBB:0.089651,STYLE:0.086828):0.013089):0.011165,TETPY:0.079158):0.012733):0.008750,(PLAFK:0.136821,(((DROME:0.018667,ARTSA:0.021658):0.006242,HUMAN:0.031416):0.030918,ENTHI:0.096304):0.012059,((PODAN:0.014754,TRIRE:0.018692):0.013881,(PUCGR:0.035030,YEAST:0.040696):0.011081):0.060281):0.006108):0.011111):0.033730):0.372402):0.000000;
```

The DIVERGE results were analyzed in the context of a 3D structure using *Pymol* software. The posterior probability (PP) cut-off we chose to select sites to mutate with EF-Tu or eEF1A were;

- a) Sites predicted to have experienced Type-1 functional divergence in which the $PP \geq 0.9$
- b) Sites predicted to have experienced Type-1 functional divergence in which the $PP \geq 0.80$
- c) Sites predicted to have experienced Type-1 functional divergence in which the $PP \geq 0.80$ + sites predicted to have experienced Type-2 functional divergence

These sites were analyzed in the context of the crystal structures of Tu and 1A, and the sites within 5 Angstroms of the nucleotide exchange factors were selected to be swapped between Tu and 1A experimentally. Using these conditions, we hypothesized that generating these mutations would knockout EF-Ts binding for EF-Tu and knockin eEF1B binding into EF-Tu. The converse applies to the knockout of eEF1B binding for eEF1A and the knockin of EF-Ts binding for eEF1A.

2.2. MATERIALS AND METHODS

Materials

Escherichia coli strains and genotypes

Tuner (DE3) pLysS Competent Cells (*Novagen*): The Tuner strain is a mutant form of K12, which contains the *lac permease* mutation (*lacZY*) that allows adjustable levels of protein expression by permitting entry of IPTG into the cells. The Tuner strain has pLysS plasmid which affects cell growth and viability by stabilizing recombinants encoding target proteins. **Genotype:** $F^- ompT hsdS_B (r_B^- m_B^-) gal dcm lacYI(DE3) pLysS (Cam^R)$. Antibiotic resistance: Chloramphenicol and Carbenicilin.

NovaBlue Singles Competent Cells (*Novagen*): The NovaBlue cell is a K-12 strain of *E.coli*, which provides high yields of plasmid DNA. These competent cells are used for plasmid transformation. Antibiotic resistance: Carbenicilin. **Genotype:** $endA1 hsdR17 (r_{K12}^- m_{K12}^+) supE44 thi-1 recA1 gyrA96 relA1 lac F[proA^+ B^+ lacI^q \Delta MI5::Tn10] (Tet^R)$

C41 (DE3) pLysS Chemically Competent Cells (*Lucigen*): These cells are usually preferred due to their tolerance to toxic proteins. In this study, the purpose of using these cells was to express recombinant proteins that usually form inclusion bodies in the Tuner cells. Also, these cells effected protein purity for some variants. **Antibiotic resistance:** Chloramphenicol. **Genotype:** $F^- ompT hsdSB (r_B^- m_B^-) gal dcm (DE3) pLysS (CmR)$.

Vectors

pET-15b (*Novagen*): This vector contains a T7 promoter and a N-terminal His-Tag which is followed by a thrombin cleavage site. This vector was used for cloning and expression of nucleotide exchange factors.

pET-21a (*Novagen*): This vector contains a T7 promoter and a C-terminal His-Tag sequence. In this study, this vector used for cloning and expression of EF-Tu and eEF1A variants.

pET-41a (*Novagen*): This vector contains T7 promoter and both GST-Tag and His-Tag sequences. Also, the vector has thrombin and enterokinase cleavage sites. **Antibiotic resistance:** Kanamycin. Purpose of using this vector was increase solubility.

pET-43.1a (*Novagen*): This vector contains T7 promoter and 491 amino acids Nus-Tag protein. The vector also contains His-Tag and S-tag sequences which are followed by thrombin and enterokinase cleavage sites, respectively. The purpose of using this vector was to increase solubility due to the large Nus-Tag sequence.

General Medias and Buffers

LB (Luria-Bertani) Media: 0.5% yeast extract, 1 % tryptone, 0.5 % NaCl, and 1.5% agar for plates.

YETM: 0.5% yeast extract, 1 % tryptone, 1% magnesium Sulfate heptahydrate ($\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$), and 1.5% agar for plates. Adjust to pH 7.5 with KOH.

TFB1: 30mM potassium acetate (KOAc), 100mM Rubidium chloride (RbCl), 10mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 50mM $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, and 15% Glycerol. Adjust to pH 5.8 with 0.2 M acetic acid.

TFB2: 10mM MOPS, 75 mM CaCl₂-2H₂O, 10mM RbCl, and 15% Glycerol. Adjust to pH 6.5 with KOH.

YPD medium and plates: 1% yeast extract, 2% peptone, 2% dextrose (D-glucose), and 2% agar for plates.

10XTBE Buffer: 10.8% Tris base, 5.5% Boric acid, and 0.93% EDTA

10XSDS (Sodium Dodecyl Sulfate) Buffer: 3.3 % Tris Base, 14.4 % Glycine, and 1% SDS

Fairbanks Gel Staining/Destaining Solutions: A (0.05% Coomassie, 25% isopropanol, 10% acetic acid). B (0.005% Coomassie, 10% isopropanol, 10% acetic acid). C (0.002% Coomassie, 10% acetic acid). D (10% acetic acid).

Genes and Primers

The genes that were used in this study were synthesized by *DNA2.0 Inc.* and *Genewiz Inc.* Synthetic PCR primers were purchased from *Integrated DNA Technologies (IDT) Inc.* Primers were used for subcloning into different vector. The genes and primers were diluted up to 100 µM and stored at -20 °C.

Methods

Preparation of Competent Cells

The purchased frozen cells were streaked onto a YETM plate and incubated at 37 °C. A single colony from YETM plate was inoculated into 5 mL of YETM medium and incubated for overnight. The overnight culture was diluted into 250 mL YETM medium and incubated for 2 hr with agitation, followed by additional 15 min incubation on ice. The cells were collected by centrifugation at 2,000 rpm at 4 °C and resuspended in 10 mL of TFB1 buffer. After 5 min

incubation on ice, the centrifugation step was repeated and the pellet was resuspended in 2 mL of TFB2 buffer and incubated on ice for 15 min. The competent cells were stored at -80 °C in microcentrifuge tubes.

Subcloning EF-Ts and eEF1B

Initially, we received our EF-Ts and eEF1B genes in the pET21a vector, which contains a His-Tag sequence at the carboxy terminus of the gene. We wanted to remove His-Tag sequence with a thrombin cleavage site. The following primers were used for subcloning from pET21a to pET15b.

Amplifying forward primer: 5'TAATACGACTCACTATAGGGGAAT3'

Amplifying EF-Ts reverse primer: 5'GTGCTCGAGTTAGCTCTGCTT3'

Amplifying eEF1B reverse primer: 5'GTGCTCGAGTTACAATTTCTGCAT3'

Basically, pET21a was used as template and the target region was amplified by PCR under standard conditions. After gel purification of amplified-DNA, both insert and pET15b vector were double digested with NdeI and XhoI (NEB) restriction enzymes for overnight at 37 °C. Standard ligation and chemical transformation protocols were applied. The enzymes used for ligation was purchased from *New England Biolabs* (NEB).

Recombination PCR

PCR reactions were performed with *Taq* polymerase and optimized for PCR yield unless otherwise specified. The conditions are summarized in Table 2.1.

Table 2.1: General PCR cycling parameters for subcloning. Annealing temperature vary depending on variant and primers.

Component	Final Concentration	Volume/25μL rxn
-----------	---------------------	-----------------

Water		14.25 μ L
5X Go Taq colorless Buffer	5X	5 μ L
dNTPs	10mM	0.5 μ L
Amplify forward primer	10 μ M	2 μ L
Amplify reverse primer	10 μ M	2 μ L
Template plasmid DNA	20 ng/ μ L	1 μ L
Go Taq polymerase	5U/ μ L	0.25 μ L

Step	Time	Temperature	Cycles
Initial Denaturation	3 min	95 °C	1X
Denaturation	30 sec	95 °C	
Annealing	30 sec	52 °C	34X
Extension	90 sec	72 °C	
Final Extension	6 min	72 °C	1X

QIAgen Gel Extraction Protocol

Fragments of DNA generated by PCR reactions were separated using standard DNA electrophoresis (0.8 % agarose gel). DNA bands corresponding to desired products were identified using a UV transilluminator and bands were excised from EtBr-stained gels using a scalpel. Separation of DNA from gel was achieved using the QIAgen Gel Extraction Kit and protocols supplied by the manufacturer (*QIAgen*). Basically, the gel slice was weighted in a 1.5 mL microcentrifuge tube and mixed with 3 volumes of QG buffer, containing guanidine thiocyanate. Samples were incubated at 48 °C until the gel slice had completely dissolved and the optimal pH for DNA binding to the spin column was adjusted by using 3 M sodium acetate (pH 5.0). To increase the yield of DNA fragments, 1 gel volume of isopropanol was added to the

sample and mixed. The sample was applied to the column, provided by manufacturer, and centrifuged. The column was washed with PE buffer, containing ethanol, and the DNA was eluted with 50 μ L sterile molecular grade water (PH 7.5). To increase DNA concentration, the tube incubated with elution water for 3-5 min at room temperature prior to centrifugation. The eluted DNA was stored at -20°C .

Chemical Transformation

The frozen competent cells were placed on ice to thaw. 10 μ L of competent cells were mixed with 50 ng of plasmid DNA and incubated on ice to ensure an even distribution of DNA. Cells were heat-shocked at 42°C water bath for 45 sec, following 2 min incubation on ice. 250 μ L of LB media was added to the cells and placed on shaker for 1 hr at 37°C . 75 μ L of transformation mixture was plated onto LB agar plates with a plasmid specific selective antibiotic and incubated at 37°C for overnight.

Analyzing Transformants

Several single white colonies were taken from the selective plate and cultured individually in 4 mL LB media containing suitable antibiotic. The culture was placed in a 37°C incubator with 250 rpm shaking for overnight.

Plasmid DNA Purification

The plasmid DNA was isolated using QIAprep Spin Miniprep Kit from *QIAGEN*. The bacterial cells were transferred to a microcentrifuge tube and collected by centrifugation* at room temperature. The pellet was resuspended in buffer P1 (lysis buffer and RNase-A), buffer P2 (contains sodium hydroxide), and buffer N3 (contains guanidine hydrochloride, acetic acid), respectively. After cells lysate was clarified by centrifugation, the supernatant was loaded to a

QIAprep spin column and washed with buffer PE. The plasmid DNA elution was collected in a 1.5 mL sterile microcentrifuge tube with 50 μ L sterile water. The concentration of the samples was determined using Nanodrop 1000 spectrophotometer, *Thermo Scientific Inc.*

*All centrifugation steps are carried out at 13,000 rpm ($\sim 17,900 \times g$) in a conventional, table-top microcentrifuge and all steps carried out at room temperature.

Sequencing

To confirm whether the DNA plasmids were cloned in the correct orientation, the samples were sequenced by *Genewiz Inc.* The samples were diluted to 50 ng/ μ L by using sterile water. For each reaction, 10 μ L of the plasmid DNA at 50 ng/ μ L provided in a labeled 0.2 mL PCR tube and cap.

Expression and Purification of EF-Tu, EF-Ts, and eEF1B

A single colony was inoculated into 3 mL of LB media and incubated for overnight at 37 $^{\circ}$ C. The overnight culture of Tuner cells was diluted into a flask containing 250 mL of LB media, 100 μ g/mL carbenicillin, and 50 μ g/mL chloramphenicol. The cells were grown at 37 $^{\circ}$ C to a density of 0.6 at A_{600} and then induced to a final concentration at 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG). The culture was incubated at 37 $^{\circ}$ C with agitating at 250 rpm for 4 hr. The cells were collected by centrifugation at $4,500 \times g$ for 20 min. The wet weight of the pellet was determined and stored at -80 $^{\circ}$ C until was resumed purification.

The freezer cells were thawed at room temperature for 15 min and resuspended in BugBuster Protein Extraction Reagent (*Novagen*). The cell pellet was completely resuspended in BugBuster reagent by pipetting and gentle vortexing, using 5 mL BugBuster per gram of wet cell pellet. In most cases, 1 μ L (25 units) of Benzonase (*Novagen*) was added per mL of BugBuster reagent to digest DNA and reduce sample viscosity. The cell suspension was incubated on a

shaking platform for 30 min at room temperature. The insoluble cell debris was removed by centrifugation at $10,000 \times g$ for 30 min at 4 °C.

The collected supernatant was transferred to a clean tube and the pellet was saved to analyze inclusion bodies. A Ni-NTA Spin Column (*QIAGEN*) was equilibrated with Lysis Buffer (50 mM NaH_2PO_4 , 500 mM NaCl, 5 mM MgCl_2 , and 5 mM imidazole, pH 7.6) and centrifuged for 2 min at $500 \times g$. The cleared lysate was loaded onto the pre-equilibrated Ni-NTA spin column and centrifuged. The column was washed three-five times with Wash Buffer (50 mM NaH_2PO_4 , 500 mM NaCl, 5 mM MgCl_2 , and 50 mM imidazole, pH 7.6). The protein was eluted twice with Elution Buffer (50 mM NaH_2PO_4 , 500 mM NaCl, 5 mM MgCl_2 , and 500 mM imidazole, pH 7.6). In all steps, the flow-through was saved for analysis by SDS-PAGE to check the stringency of the conditions. The purified protein was dialyzed against 50 mM Tris-HCl (pH 7.8), 100 mM KCl, 1 mM DTT, and 3 mM MgCl_2 .

Expression and Purification of EF-Tu Variants, eEF1A, and its Variants

A single colony was inoculated into 4 mL of LB media and incubated for overnight at 37 °C. The overnight culture diluted in a flask containing 250 mL of LB medium, 100 µg/mL carbenicillin, and 50 µg/mL chloramphenicol. The cells were grown at 37 °C to a density of 0.6 at A_{600} and then induced with 0.5-1 mM IPTG. The culture was incubated at 37 °C and agitated at 250 rpm for 6 hr. The cells were collected by centrifugation at $5,000 \times g$ for 15 min. The wet weights of the pellet were determined and immediately purified or stored at -80 °C until needed.

The freezer cells were thawed at room temperature for 15 min and resuspended in 8 M Urea; 100 mM NaH_2PO_4 ; 5 mM MgCl_2 ; and 10 mM Tris-HCl (pH 8.0). The cell pellet was completely resuspended in buffer by pipetting and gentle vortexing. Cell lysis was achieved by

sonication and the suspension was incubated on a shaking platform for 2 hr at room temperature. The cell debris was removed by centrifugation at $10,000 \times g$ for 30 min at room temperature.

The supernatant was transferred to a clean tube and Ni-NTA Spin Columns (*Qiagen*) was equilibrated with Lysis Buffer (8 M Urea; 10 mM NaH_2PO_4 ; 5 mM MgCl_2 ; and 10 M Tris-HCl (pH 8.0)) and centrifuged for 2 min at $600 \times g$. The cleared lysate was loaded onto the pre-equilibrated Ni-NTA spin column and centrifuged. The column was washed three times with 8 M Urea; 100 mM NaH_2PO_4 ; 5 mM MgCl_2 ; and 10 mM Tris-HCl; 50 mM imidazole, pH 8.0. The elution was performed with 8 M Urea; 100 mM NaH_2PO_4 ; 5 mM MgCl_2 ; 10 mM Tris-HCl; and 300 mM imidazole, pH 8.0. In all steps, the flow-through was saved for analysis by SDS-PAGE to check the stringency of the conditions.

Refolding Denatured-Purified Proteins

Denatured purified-protein was diluted two-fold in buffer of 50 mM Tris, 20 mM NaCl, 100 mM KCl (pH=8.2), and then stepwise dialyzed against 2 M urea, 50 mM Tris-HCl, 20 mM NaCl, 100 mM KCl, (pH=8.2); 1 M urea, 50 mM Tris-HCl, 20 mM NaCl, 100 mM KCl, (pH=8.2); 0.5 M urea, 50 mM Tris-HCl, 20 mM NaCl, 100 mM KCl, (pH=8.2), respectively, by using Molecular weight cut-off (MWCO) dialysis cassette (*Pierce*).

Removing His-tag from EF-Ts and eEF1B

The His-tag was removed by using the Thrombin Cleavage Capture Kit (*Novagen*). Thrombin was diluted in 1:25 in Thrombin Dilution Buffer (50 mM sodium citrate, pH 6.5, 200 mM NaCl, 0.1% PEG-8000, 50% glycerol) and mixed with target protein and 10X Thrombin Cleavage Buffer (200 mM Tris-HCl, 1.5 M NaCl, 25 mM CaCl_2 (pH=8.4)). The mixture was incubated at 25 °C with agitation for 20 hr. Reactions were stopped with either protease inhibitor

complex or Streptavidin Agarose (50 % slurry in phosphate buffer, 0.02 % sodium azide (pH=7.5)) and incubated for an additional hr. The protein that still has His-tag was removed via loading on pre-equilibrated His-tag spin column. The flow through not containing His-tag was collected for analysis.

Protein Quantification

All the proteins used in binding assays were quantified by performing Bradford protein assay. Each sample was measured in triplicate. Basically, 1 mL of Bradford solution (*Bio-Rad*) was incubated at room temperature for 30 min and followed by the addition of 20 μ L of sample protein. After 5 min incubation at room temperature, samples were transferred to a cuvette and absorption at 595 nm was plotted against a bovine gamma globulin concentration curve (*Cary 50 UV-Visible Spectrophotometer, Varian*).

Pull-Down Assay

His-tagged eEF1A and EF-Tu variants were mixed with either eukaryotic (eEF1B) or bacterial (EF-Ts) nucleotide exchange factors (without His-tag) at a ratio of 1:1.5 and incubated at room temperature for 4 hr in buffer of 55 mM Tris-HCl (pH 7.8), 130 mM KCl, 20 mM NaCl, and 2 mM EDTA. Before loading on the Ni^{+2} -NTA column, samples were diluted two-fold in incubation buffer without EDTA. The dilution was done to prevent interference of EDTA with the column. The column was washed twice with buffer A containing 55 mM Tris-HCl (pH 8.2), 20 mM NaCl, 10 mM KCl, 500 mM urea, and 10 mM imidazole, to remove nonspecific proteins. Finally, samples were eluted with buffer A containing 55 mM Tris-HCl (pH 8.2), 20 mM NaCl, 10 mM KCl, 500 mM urea, and 300mM imidazole and analyzed by SDS-PAGE.

Native Gel Electrophoresis

Native gel electrophoresis was performed at 4 °C (250 V, 1.5 hr) on 10% Tris-HCl polyacrylamide gel. The running buffer used contained 8 mM Tricine (pH 8.2) and 1 mM EDTA. To analyze the interactions between EF-Tu and EF-Ts, these proteins were mixed in a buffer that contained 50mM Tris-HCl (pH 7.8), 70mM KCl, 5mM EDTA, 1mM DTT, and 5% glycerol. After 2 hr incubation at room temperature, the mixture was mixed with a half volume of native gel sample buffer and then load on the native gel. The gel was stained with standard gel staining/destaining solutions (*Fairbanks*).

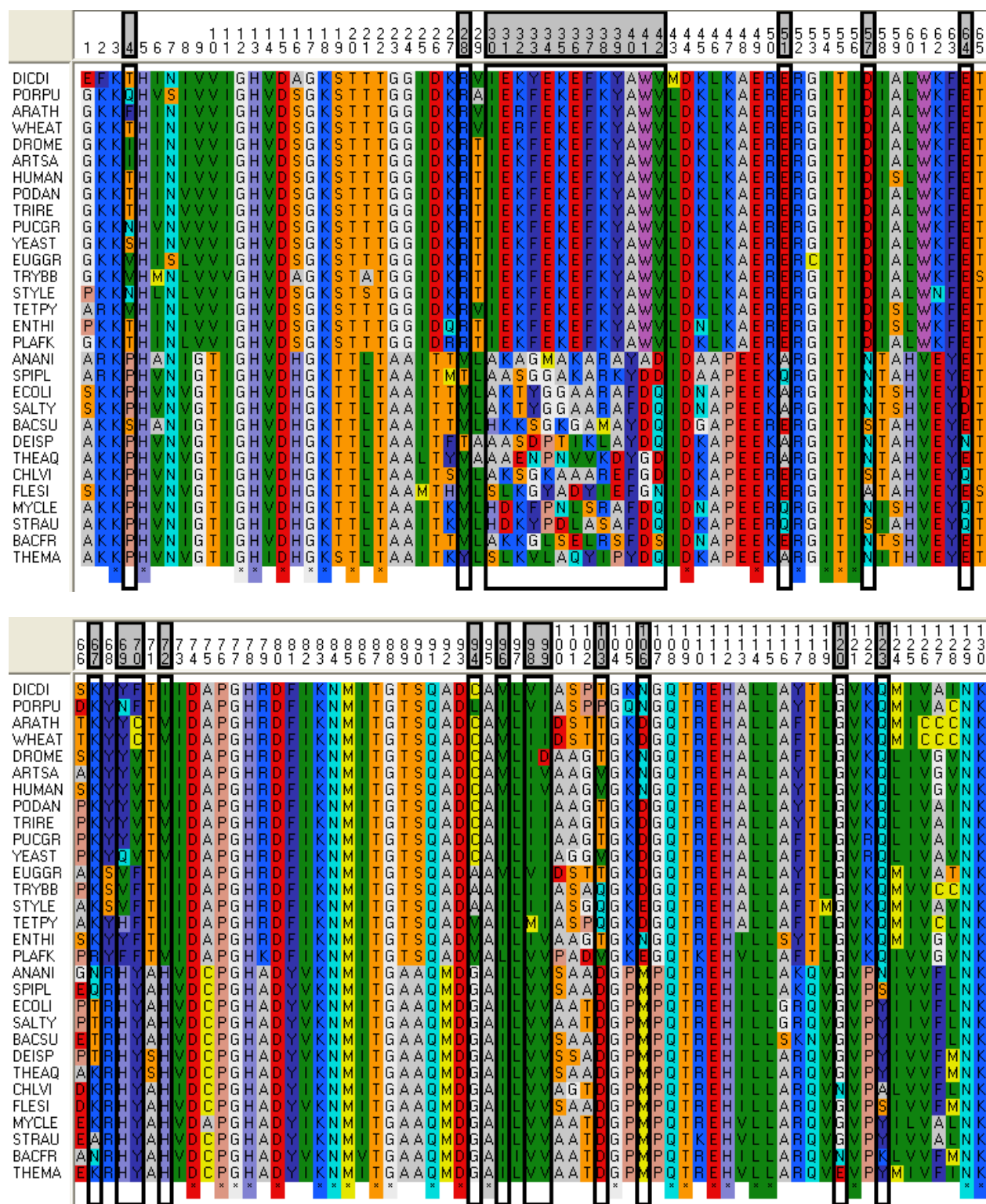
2.3. RESULTS AND DISCUSSION

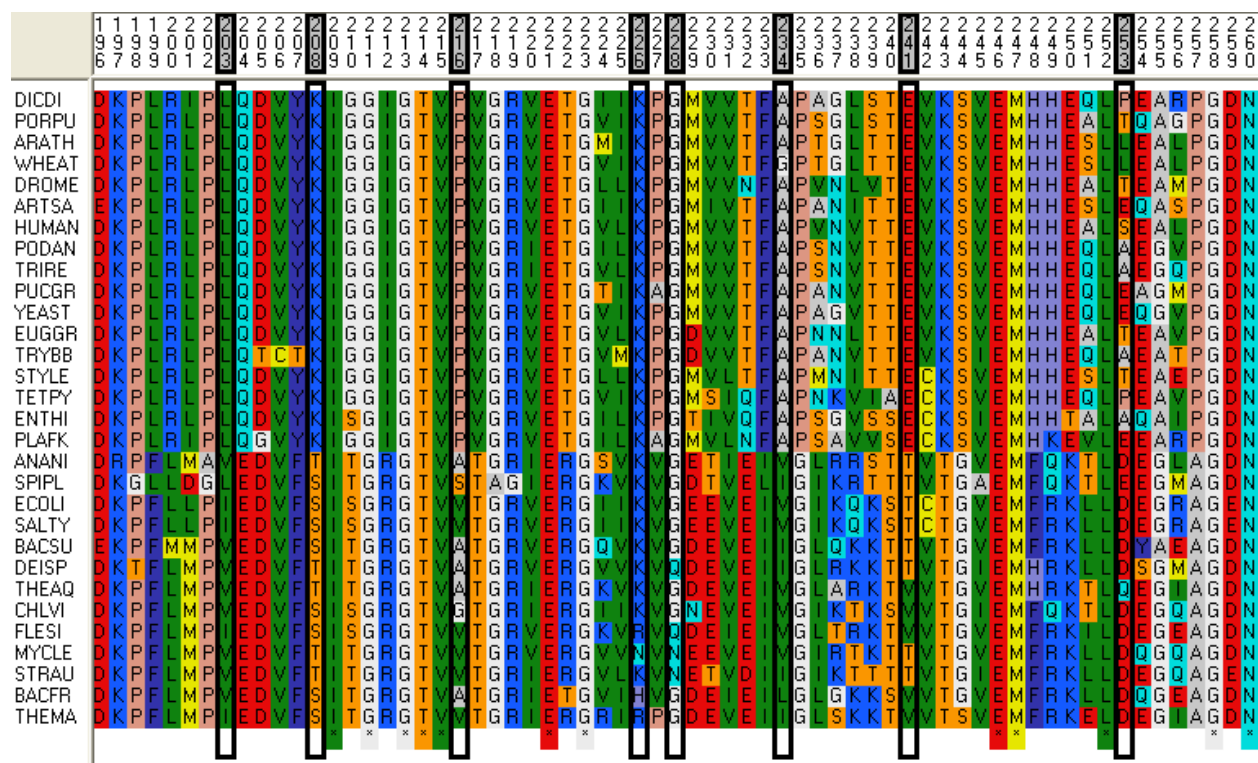
Despite intense efforts and interest in understanding patterns of functional divergence, no one has experimentally validated these concepts to date. To address this, we have mutated specific sites identified by our evolutionary analyses as being implicated in functional divergence between EF-Tu and eEF1A and their abilities to bind their respective nucleotide exchange factors.

Detecting Functionally Divergent Sequences

We performed DIVERGE software analysis to detect functional divergence among EF protein family members, based on site-specific rate shifts. Posterior probabilities of whether a site may have experienced type-I or type-II functional divergence guided us in our selection of amino acid residues believed to be responsible for functional divergence among this protein family. We used EFs sequences from 13 species in bacteria and 17 species of eukaryotes. Based on the DIVERGE analysis, residues predicted to have experienced type-I functional divergence between Tu and 1A homologs are highlighted in Figure 2.9. For example, positions 31, 34, 36, 51, 64 etc. are occupied by a conserved glutamate in eukaryotes. Conversely, these same positions are occupied by various amino acids in bacteria and thus those sites are considered to be rapidly evolving across time. Other examples of functional divergence based on the DIVERGE analysis are positions 9 and 10 which are occupied by a conserved hydrophobic valine in eukaryotes while the same positions are occupied by other conserved but different amino acid glycine (hydrophobic) and threonine (hydrophilic), respectively. This distribution is the signature of type-II functional divergence between these two domains of life. Some sites (i.e. 3, 5, 12, 13, 15, etc) that are thought to be critical in functions that are shared by all members of

the EF gene family (e.g., shuttling aa-tRNAs) are conserved in both domains of life, eukaryotes and bacteria.





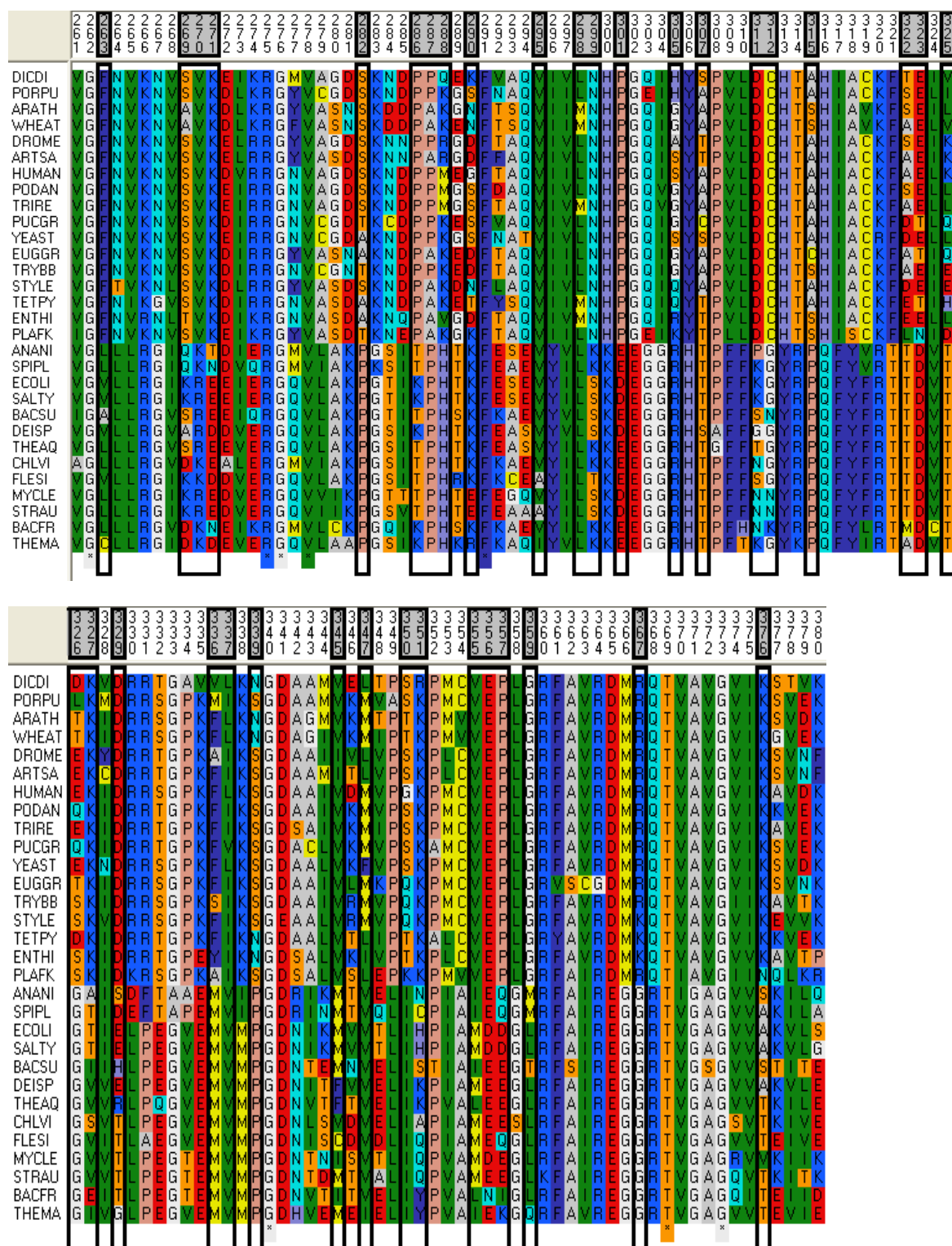


Figure 2.9: Sequence alignment of Eukaryotic and Bacterial EFs. Boxed positions represent type-I functional divergence. Cut-off for posterior probability is 0.9. We used EFs sequences from 13 different species in bacteria and 17 species from eukaryotes. Type-II sites are not highlighted but they are easily identifiable from patterns of amino acid replacements (e.g., position 371).

Detecting mutated sites

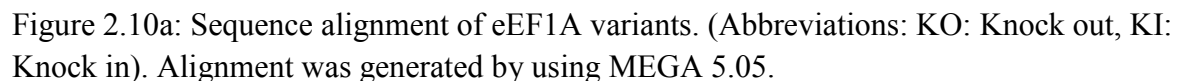
Pymol software was used to determine the distribution of functionally divergent sites across the three-dimensional structures of the proteins. In Pymol, we aligned the 1EFU and 1F60 structures to identify sites that were within 5 Angstroms of the opposite nucleotide exchange factor (e.g., sites on EF-Tu that would be within 5 Angstroms if eEF1B were to bind EF-Tu, and conversely for eEF1A). Sites to mutate were selected based on the DIVERGE analysis using parameters and cutoffs summarized in table 2.2.

Table 2.2: Considered statistical support and parameters to generate mutant variants. All cut-offs within 5 Angstroms of respective nucleotide exchange factor.

Eukaryotic background (eEF1A)		
Substitution in Nucleotide Exchange Factor	Variant name	Conditions and Cutoffs
Knockout eEF1B	E1A	Type 1 Divergence with $PP \geq 0.9$
	E1B	Type 1 Divergence with $PP \geq 0.80$
	E1C	Type 1 Divergence with $PP \geq 0.80$ + Type 2
Knockin EF-Ts	E2A	Type 1 Divergence with $PP \geq 0.9$
	E2B	Type 1 Divergence with $PP \geq 0.80$
	E2C	Type 1 Divergence with $PP \geq 0.80$ + Type 2
Knockout eEF1B/Knockin EF-Ts	E3A	Type 1 Divergence with $PP \geq 0.9$
	E3B	Type 1 Divergence with $PP \geq 0.80$
	E3C	Type 1 Divergence with $PP \geq 0.80$ + Type 2
Bacterial background (EF-Tu)		
Substitution in Nucleotide Exchange Factor	Variant name	Conditions and Cutoffs
Knockout EF-Ts	B1A	Type 1 Divergence with $PP \geq 0.9$
	B1B	Type 1 Divergence with $PP \geq 0.80$

	B1C	Type 1 Divergence with $PP \geq 0.80$ + Type 2
Knockin eEF1B	B2A	Type 1 Divergence with $PP \geq 0.9$
	B2B	Type 1 Divergence with $PP \geq 0.80$
	B2C	Type 1 Divergence with $PP \geq 0.80$ + Type 2
Knockout EF-Ts / Knockin eEF1B	B3A	Type 1 Divergence with $PP \geq 0.9$
	B3B	Type 1 Divergence with $PP \geq 0.80$
	B3C	Type 1 Divergence with $PP \geq 0.80$ + Type 2

We generated mutant variants of EF-Tu and eEF1A based on the combination of DIVERGE results and location on protein structure. The number of mutated sites within a variant varied from 4 to 51 sites depending on the analysis and statistical cutoff considered.



	10	20	30	40	50	60	70	80	90	100
EF-Tu	MSKEKFERTKPHVNVGTIGHVDHGKTTLTAAITTVLAKTYGGAARAFDGDIDNAPEEKARGITINTSHVEYDTPTRHYAVVDCPGHADYVKNMITGAAQMD									
B1A KO-Ts			K		EFKYAWV		RE	D		
B1B KI-1B			K	E	EFKYAWVL	K	RE	D	I	
B1C KOTs/KI1B		V	G	DK	E	EFKYAWVL	K	KA	RE	D
B2A KO-Ts					Q					
B2B KI-1B					Q		I	K	I	
B2C KOTs/KI1B					Q		I	L	K	I
B3A KO-Ts			K		EFKYAWV		RE	D		
B3B KI-1B			K	E	EFKYAWVL	K	RE	DI	K	I
B3C KOTs/KI1B		V	G	DK	E	EFKYAWVL	K	KA	RE	DI
	110	120	130	140	150	160	170	180	190	200
EF-Tu	GAILVVAATDGPMPQTRREHILLGRQVGVPIIIVFLNKCMDVDEELLELVEMEVRRELLSQYDFPGDDTPIVRGSALKALEGDAEWEAKILELAGFLDSYI									
B1A KO-Ts		GV	D		S	S			T	
B1B KI-1B		GV	KD		S	SR	I		NT	
B1C KOTs/KI1B		GV	KD		S	SR	EI		NT	
B2A KO-Ts										
B2B KI-1B										
B2C KOTs/KI1B										
B3A KO-Ts		GV	D		S	S			T	
B3B KI-1B		GV	KD		S	SR	I		NT	
B3C KOTs/KI1B		GV	KD		S	SR	EI		NT	
	210	220	230	240	250	260	270	280	290	300
EF-Tu	PEPERAIDKPFLLPIEDVFSISGRGTVVTVGRVERGIIKVGEVEIVGIKETQKSTCTGVEMFRKLLDEGRAGENGVLLRGIKREEIERGQVLAKPGTIK									
B1A KO-Ts										
B1B KI-1B										
B1C KOTs/KI1B										
B2A KO-Ts										
B2B KI-1B							E	D		
B2C KOTs/KI1B		Q					H	E	D	
B3A KO-Ts										
B3B KI-1B							E	D		
B3C KOTs/KI1B		Q					H	E	D	
	310	320	330	340	350	360	370	380	390	
EF-Tu	PHTKFESEVYILSKDEGGRHTPFFKGYRQPFYFRTTDVTGTIELPEGVEMVMPGDNIKVVTLIHPIAMDDGLRFAIREGGRTVGAGVVAKVLS									
B1A KO-Ts					FL					
B1B KI-1B			L		KFL					
B1C KOTs/KI1B			L		KFL					
B2A KO-Ts		N	P					R		
B2B KI-1B		N	P	QI				R		
B2C KOTs/KI1B		N	P	QI				RQ	A	
B3A KO-Ts		N	P			FL		R		
B3B KI-1B		N	P	QI	L	KFL		R		
B3C KOTs/KI1B		N	P	QI	L	KFL		RQ	A	

Figure 2.10b: Sequence alignment of EF-Tu variants. (Abbreviations: KO: Knock out, KI: Knock in)

Subcloning EF-Ts and eEF1B

Full length EF-Ts and eEF1B were subcloned into an N-terminal 6x-Histidine tag vector with a thrombin cleavage site, pET15b (Novagen), at the NdeI and XhoI (NEB) restriction sites. The construct was transformed into an *E. coli* Novablue strain for propagation. Desired DNA sequences were verified by sequencing (Genewiz) and then transformed into the *E. coli* Tuner strain for

expression. Carbenicillin was used to select pET15b in Nova blue strain and both carbenicillin and chloramphenicol were used to select the Tuner strain.

Expression and Purification of EF-Tu, EF-Ts, and eEF1B

To investigate the interaction of EF-Tu and eEF1A variants for nucleotide exchange factors (EF-Ts and eEF1B), proteins were over-expressed and purified by using affinity chromatography.

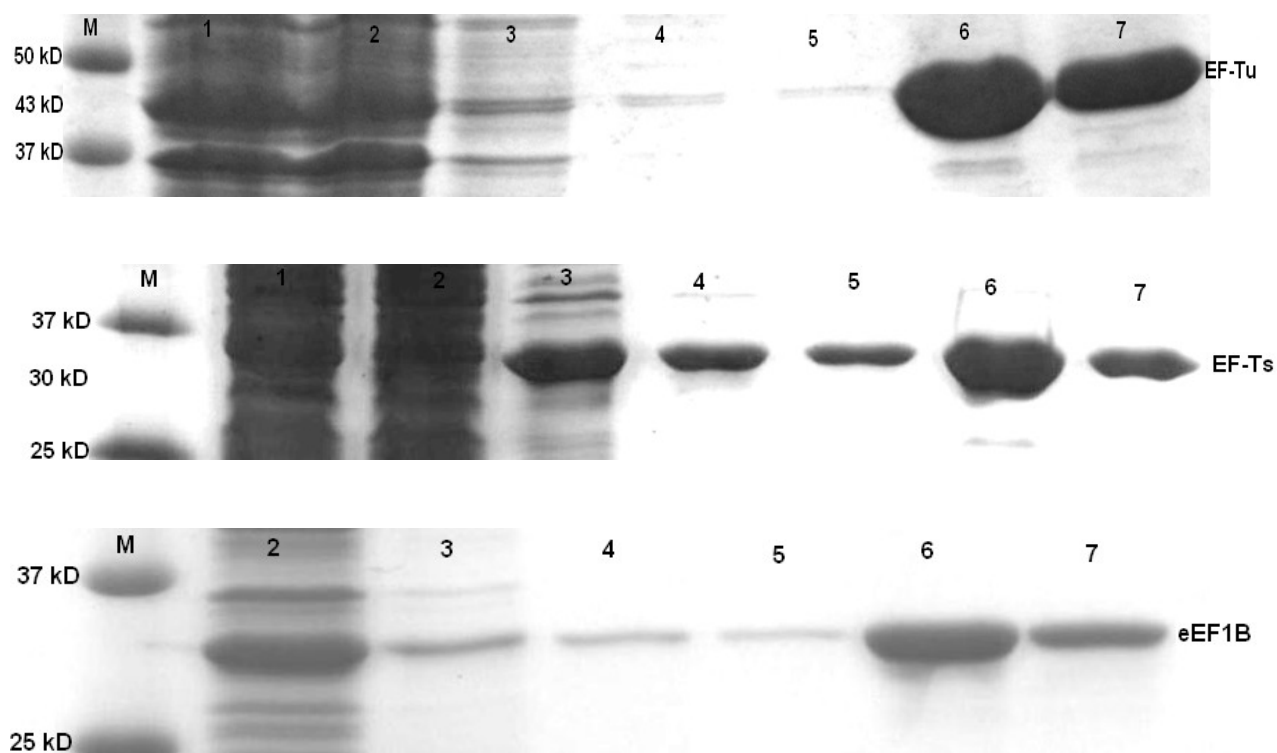


Figure 2.11: SDS-PAGE gel analysis of purified-recombinant EF-Tu, EF-Ts, and eEF1B proteins. The genes were expressed in *E. coli* and extracts were prepared under native conditions. This figure shows various purification steps using a Ni²⁺ column. Each of these proteins contain a N-terminal or C-terminal 6XHis-tag for Ni-Affinity chromatography. M represents protein standard marker (Kaleidoscope, BioRad). Lanes numbered 1 to 7 represent; crude cell lysate (1), flow-through after loading the sample onto a Ni²⁺ column (2), three-wash steps (3-5), two elution steps (6-7).

His-tagged derivative Bacterial (EF-Ts) and Eukaryotes (eEF1B) nucleotide exchange factor and EF-Tu were successfully expressed in *E. coli* and purified under standard native conditions (Figure 2.11). Expressed-proteins contain C-terminal 6XHis-tag for Ni-affinity chromatography, so purity of these proteins was achieved by using imidazole and salt gradient. From SDS-PAGE analysis the molecular weight for EF-Tu, EF-Ts, and eEF1B were roughly determined as 45 kD, 33 and 30, respectively. To remove all of unspecific proteins, a number of wash steps were conducted and an imidazole gradient was applied during elution.

Expression and Purification of EF-Tu Variants, eEF1A, and its Variants under Denaturing Conditions

Initially, a series of alternative conditions were attempted to improve expression and purification of insoluble eEF1A and the majority of the variant proteins. These conditions included varying growth and induction temperature, IPTG concentration, adjusting osmotic stress (inducing with Ethanol and Sucrose or Magnesium Chloride), expressing in different vectors (pET43.1a and pET41a) and strains (C41 and C43 cells), expressing with yeast vectors (pYES2/NT A, B, C and pPink-HC & pPink-LC) and strains (INVSc and PichiaPink strain 1, 2, 3, 4), addition of other reagents in resuspension buffer including solubilizing agents such as 1% SDS and 2% Triton, Arginine, etc. We applied all of these conditions to most of the 18 mutant EF-Tu and eEF1A variants because we were hoping that some of those proteins might respond at least to one of these experimental conditions. Despite consuming over a year to get soluble proteins, unfortunately none of the attempted modifications improved the protein solubility (data not shown). Subsequently, we solubilized the target proteins with either 8M urea or 6M guanidine

and a variety of alternate conditions were used to optimize this purification procedure including higher salt concentrations and an imidazole gradient (see materials and methods section).

Refolding Purified Protein

We have purified EF-Tu variants, eEF1A, and its variants, under denaturing conditions using 8M urea. Because denatured protein is not functional, it has to be refolded correctly before conducting structural and functional studies. We successfully refolded denatured proteins via a stepwise dilution and dialysis procedure to remove denaturing agents (see materials and methods section). Protein properties were determined by their ability to bind to nucleotide exchange factors. Results are shown in figure 2.14.

Removing His-tag from EF-Ts and eEF1B

Thrombin recognizes the consensus sequence Leu-Val-Pro-Arg-Gly-Ser and cleaves the peptide bond between Arg and Gly. This consensus sequence is utilized in pET15b vectors and allowed us to remove the His-tag from purified EF-Ts and eEF1B proteins (the His-tag was maintained on EF-Tu and eEF1A, and their variants, for purposes of immobilization during the pull-down binding assays). His-tag was removed using a Thrombin Cleavage Capture Kit (described in materials and methods). We kept some non-cleaved protein as a control and separated the cleaved protein from any non-cleaved protein by loading the reaction on a Ni-NTA column. The cleaved protein flowed through and the non-cleaved protein bound to the column and then eluted with imidazole for SDS-PAGE gel analysis (Figure 2.12). The cleaved protein was 2 KDa smaller in molecular weight than non-cleaved protein. To confirm the complete removal of His-tag protein, flow through from Ni²⁺ column was loaded onto fresh column and

eluted with imidazole. SDS-PAGE results confirm that EF-Ts (Figure 2.12) and eEF1B (data not shown) were free from His-tags.

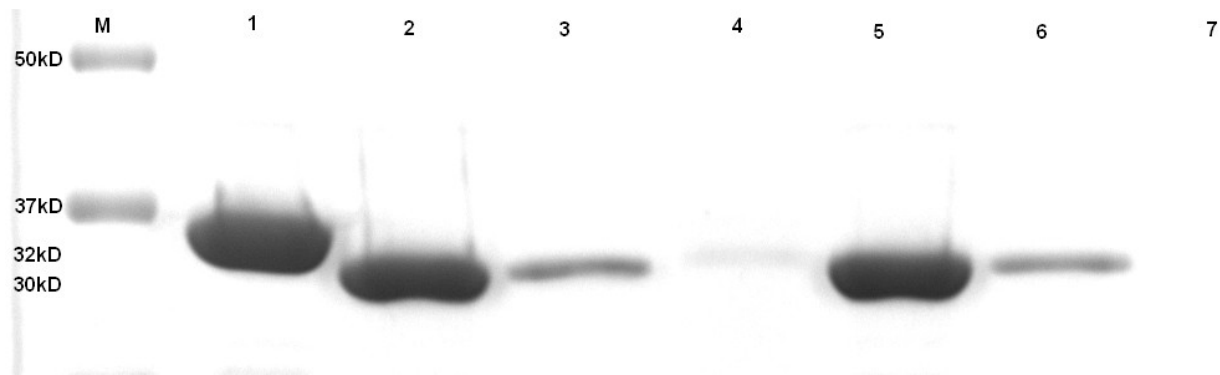


Figure 2.12: SDS-PAGE gel analysis of thrombin cleavage of EF-Ts. Purified N-terminal 6XHis-tag EF-Ts cleaved by thrombin. The cleavage reaction passed-through a Ni^{2+} column twice to remove all noncleaved EF-Ts. M represents marker to compare molecular weight of each protein (Kaleidoscope, *BioRad*). Numbers 1 to 7 represent; purified EF-Ts protein (1), first flow-through of cleaved EF-Ts after loading the sample onto a Ni^{2+} column (2), wash for nonspecific binding (3), elution (4), flow-through of number 2 (5), wash (6), elution (7).

Interaction between EF-Tu:EF-Ts and eEF1A:eEF1B

To assess whether mutant variants bind to the nucleotide exchange factors, we first looked at the interaction between wild type EF-Tu:EF-Ts and eEF1A:eEF1B. We performed native gel analysis, SDS PAGE, and size exclusion chromatography analysis to assess interactions between these pairs.

Interaction between EF-Tu:EF-Ts : Native Gel Analysis

The binding of recombinant bacterial EF-Tu to recombinant EF-Ts proteins was analyzed by native PAGE gel. In the absence of EF-Ts, a small amount of EF-Tu penetrated into the gel

while a majority stayed in the wells of the gel (Figure 2.13). This may be due to oligomerization of EF-Tu – a process known from the crystal structures of EFs. In contrast, when EF-Ts was present, complexes were formed and migrated into the gel. Perhaps, binding EF-Ts prevents oligomerization of EF-Tu. The complexes between bacterial EF-Tu and EF-Ts molecules were further confirmed by pull-down assays and size exclusion chromatography. SDS-PAGE analysis clearly revealed the presence of a EF-Tu:EF-Ts complex (Figure 2.14 and 2.15). These results served as a control while comparing binding abilities of nucleotide exchange factors to other variants of EF-Tu and eEF1A. We also tried to analyze the eEF1A:eEF1B complex on native gel but we were not successful because eEF1A has a very high isoelectric point of 9.1. Due to this high isoelectric point, penetration of eEF1A into the native gel was probably prohibited. However, the complexes between eEF1A and eEF1B molecules were confirmed by pull-down assays and size exclusion chromatography.

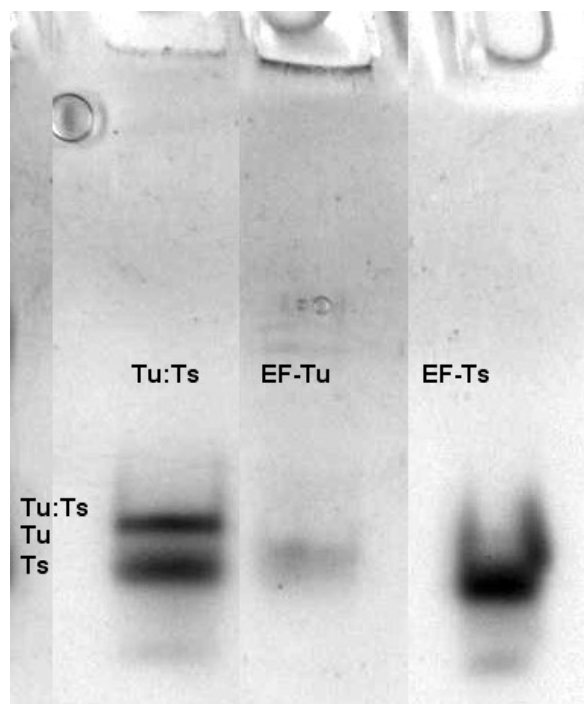


Figure 2.13: Native gel analysis of interactions between EF-Tu and EF-Ts. The ability of EF-Tu to bind to EF-Ts was analyzed by 10% Tris-HCl native gel, run for 2 hr at 250V. Tu:Ts complexes formed just above EF-Ts. Without EF-Ts, EF-Tu stayed in the well, possibly due to the oligomerization of EF-Tu molecules. The addition of EF-Ts to EF-Tu caused the EF-Tu bands to disappear and a new band containing the EF-Tu:EF-Ts complex to appear slightly above the EF-Ts band.

Interaction between EF-Tu:EF-Ts and eEF1A:eEF1B : Size Exclusion

Chromatography

The association between EF-Tu:EF-Ts and eEF1A:eEF1B were also demonstrated using a size-exclusion column. Molecules were separated based on their size. EF-Tu:EF-Ts complexes came off the column first and then followed by EF-Ts. We did not observe EF-Tu alone because most of the EF-Tu participated in EF-Tu:EF-Ts complexes thus only a small portion of EF-Tu was observed in the second peak.

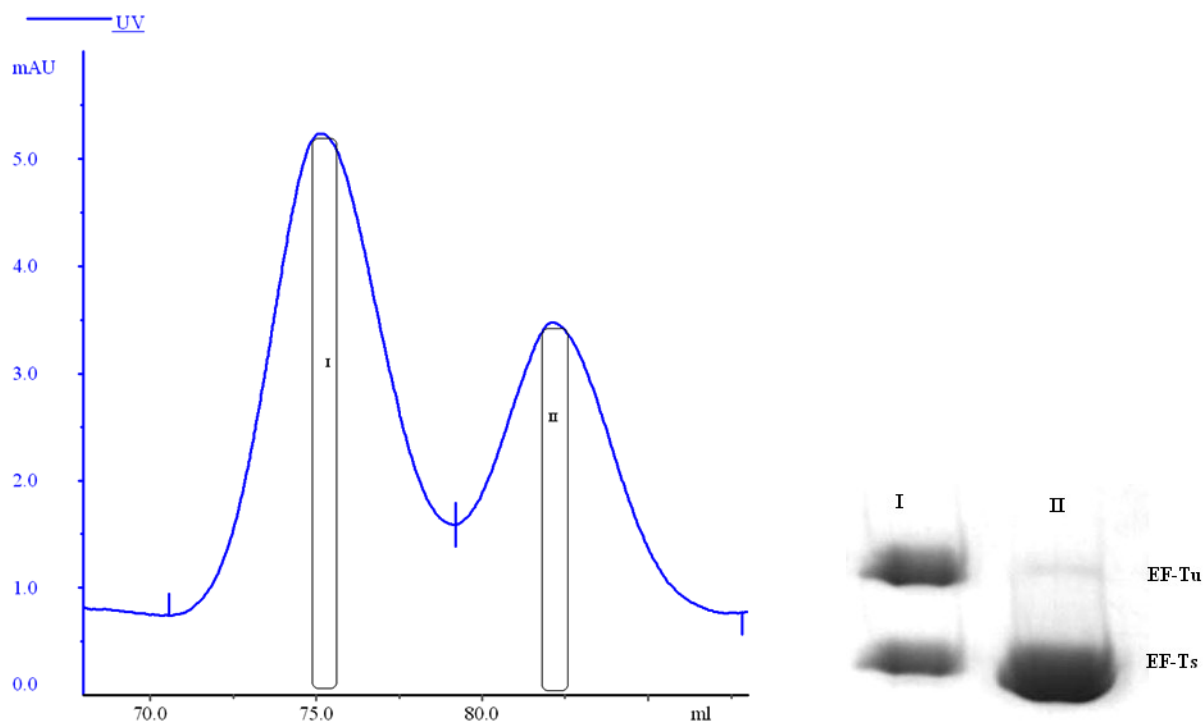


Figure 2.14: Interaction between *E. coli* EF-Tu:EF-Ts complexes analyzed by HiPrep Superdex-200 size-exclusion column chromatography. Circled area in each peak was concentrated by Trichloroacetic acid (TCA) precipitation and then analyzed by SDS-PAGE. The left peak represents the EF-Tu:EF-Ts complex (I), the right peak shows the EF-Ts molecule without a His-tag.

The complexes formed between recombinant *Saccharomyces cerevisiae* eEF1B and eEF1A efficiently revealed by size-exclusion column chromatography using a HiPrep Sephacryl S-200 column (Figure 2.15). The three peaks on the size-exclusion chromatograms for eEF1A and eEF1B were analyzed by SDS-PAGE and the presence of eEF1A:eEF1B complex was demonstrated in the middle peak (II). Separation between II and III picks were poor but SDS-PAGE analyzes revealed clear data. Some portion of eEF1A came off first, this could be because of not complete refolding of eEF1A or the interaction conditions weren't optimum.

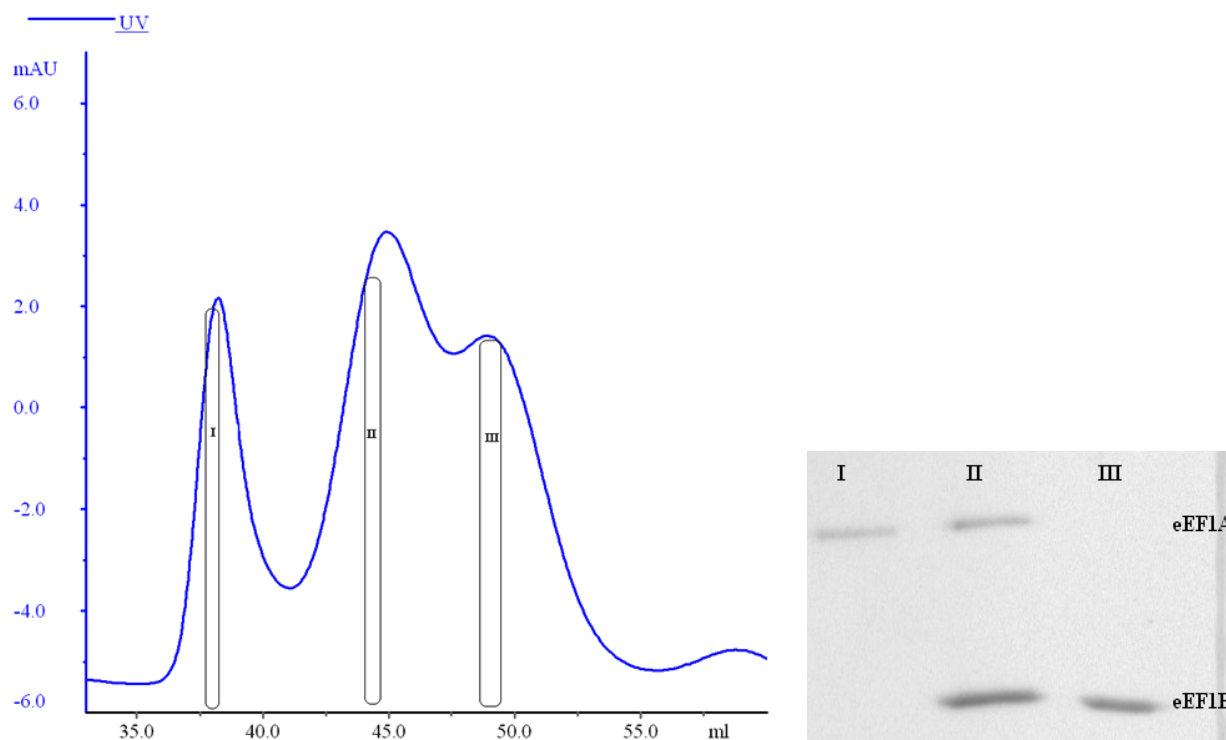


Figure 2.15: Interaction between *Saccharomyces cerevisiae* eEF1A:eEF1B complexes analyzed by HiPrep Sephacryl S-200 size-exclusion column chromatography. Circled-area in each pick was concentrated by TCA precipitation prior to analyze on SDS-PAGE. The first pick demonstrates eEF1A (I), the second pick shows eEF1A:eEF1B complex (II), and the right pick represents eEF1B molecule (III).

Interaction between EF-Tu:EF-Ts and eEF1A:eEF1B : Pull-Down Assay

Interactions between wild-type (EF-Tu:EF-Ts and eEF1A:eEF1B) and mutant variants were next monitored by performing pull-down assays. We were able to pull-down EF-Ts using a His-tagged EF-Tu as well as pull-down eEF1A using a His-tagged eEF1B (Figure 2.16). When we loaded EF-Tu or eEF1A onto the column, nearly all of the protein bound to the column based on high purity and the presence of a His-tag. However, when EF-Ts or eEF1B was loaded onto the column, nearly 50% of the protein came off with the flow-through because these proteins lack a His-tag and the portion that stayed in the column was the portion that bound to EF-Tu or

eEF1A. EF-Tu:EF-Ts complexes were eluted with a high amount of imidazole. The same experiment was performed for eEF1A:eEF1B complexes and their variants.

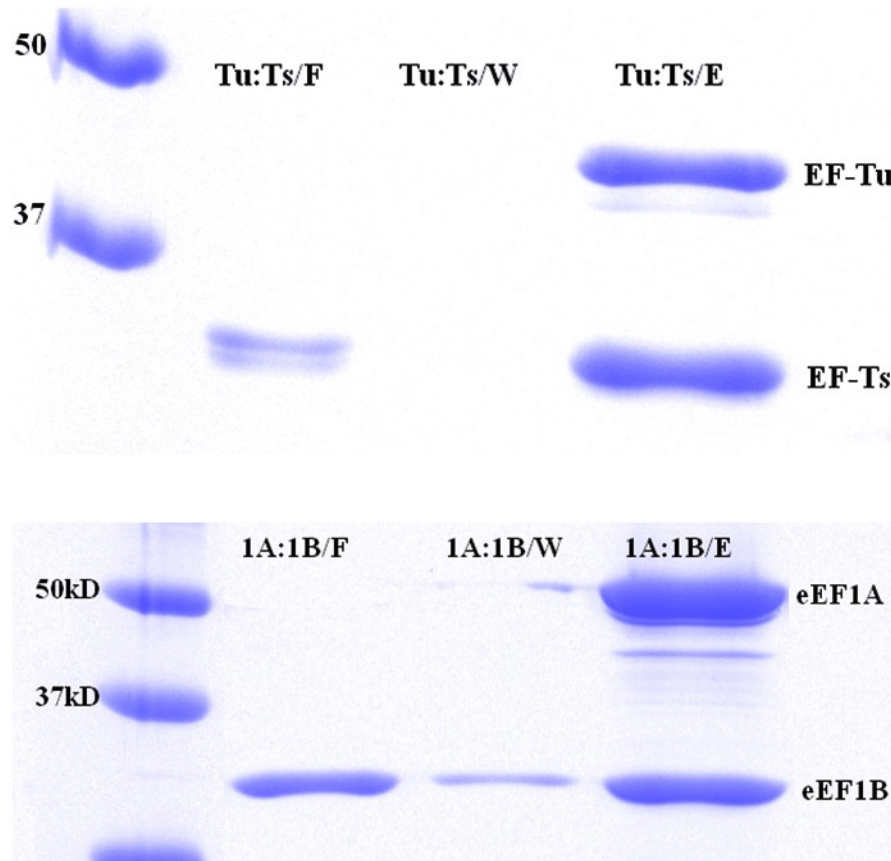


Figure 2.16: SDS-PAGE analysis of binding interactions between EF-Tu:EF-Ts and eEF1A:eEF1B. His-tagged purified EF-Tu incubated with EF-Ts (no His-Tag) and passed through a Ni^{2+} column. Non-specific interactions were removed by washing the column and then the EF-Tu:EF-Ts complex was eluted with a high amount of imidazole. The same experiment was performed for the eEF1A:eEF1B complex.

Binding ability of EF-Tu Variants to EF-Ts or eEF1B

Binding abilities of EF-Tu variants were determined by pull-down assays. Binding ability EF-Tu was analyzed in three independent steps; knock-out EF-Ts, knock-in eEF1B, and knock-out EF-Ts + knock-in eEF1B. Each step was based on three statistical parameters and structural

analysis. Therefore we had nine variants in total for each domain of life to be analyzed. Parameters are summarized in table 2.2. First we analyzed knock-out of EF-Ts from EF-Tu (based on heterotachous sites) and monitored the binding ability of EF-Tu to the nucleotide exchange factors (Figure 2.17A). EF-Ts was mostly knocked out from EF-Tu and some small portion of EF-Ts remained bound, but the bands get thinner if we compare this to the interaction between wild-type EF-Tu:EF-Ts. EF-Tu was also able to bind to eEF1B for these variants despite the fact that we did not engineer these variants to bind eEF1B. Such an intrinsic property is probably dictated by sites that have overlapping functions in EF-Tu and eEF1A in their abilities to bind their respective nucleotide exchange factors.

In the second step, we determined the ability to knock-in binding of eEF1B into EF-Tu. In this case, EF-Tu sites in proximity to eEF1B according to the overlay of the EF-Tu/eEF1B superimposition of 3D-structures and having statistically significant predictions according to DIVERGE were replaced with residues from the eukaryotic eEF1A. The amount of eEF1B that interacts with EF-Tu variants increased relative to the amount of eEF1B that can bind to wild-type EF-Tu. Interaction between EF-Tu:EF-Ts relatively decreased while compare to the wild-type interaction of EF-Tu:EF-Ts. Reduction in EF-Ts binding could be because of some mutations to knock-in eEF1B interfered with EF-Ts binding sites (Figure 2.17B).

In the third step, we replaced EF-Tu sites with eukaryotic background to create both knock-out EF-Ts and knock-in eEF1B. The binding ability of eEF1B to EF-Tu variants had gradually increase in comparison of the first and second steps. The small amount of EF-Ts remained bound to EF-Tu (Figure 2.17C). These results indicate that site-specific rate shift is an important pattern of functional diversification throughout evolutionary history.

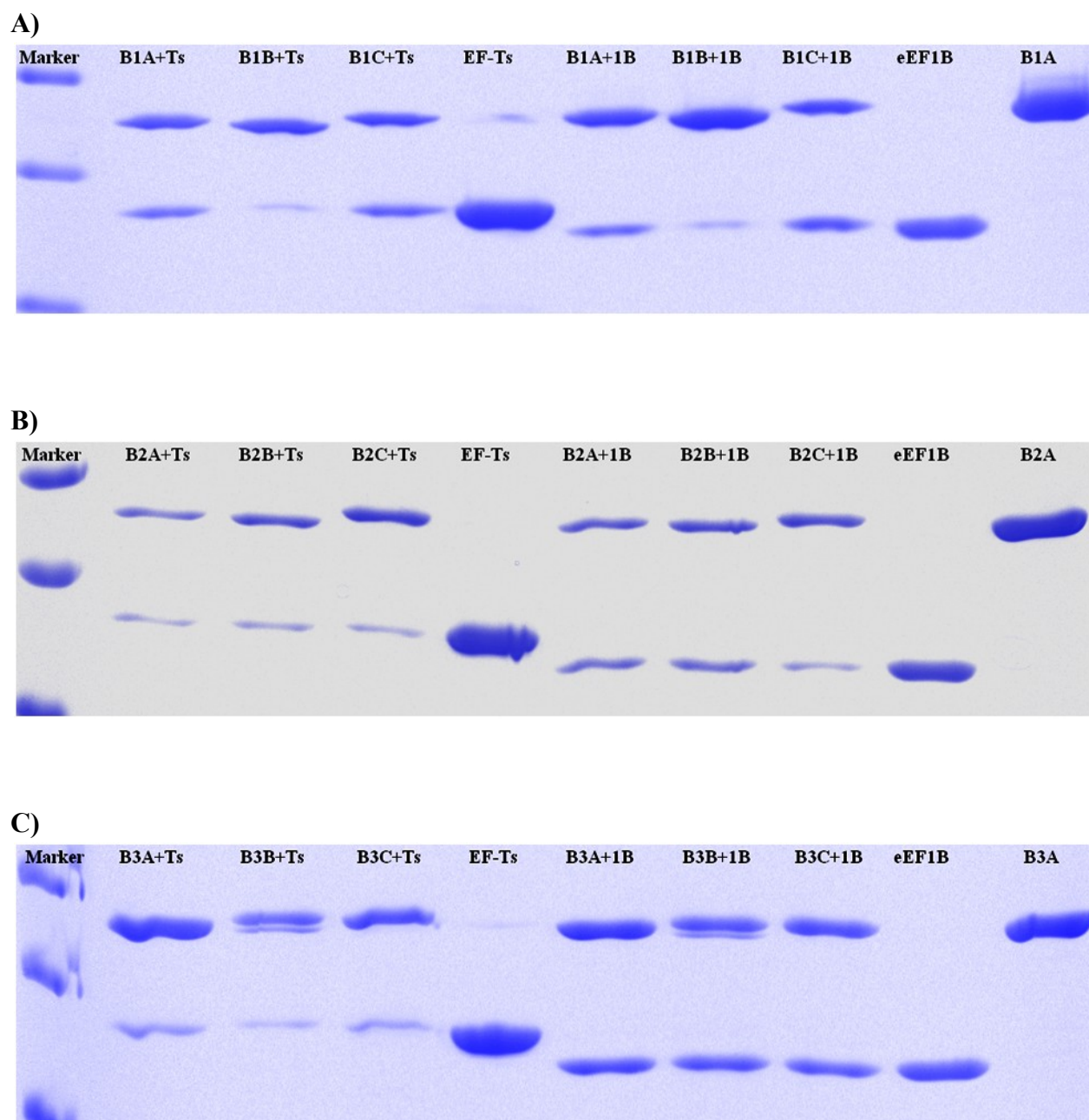


Figure 2.17. SDS-PAGE gel analysis of binding ability of nucleotide exchange factors to EF-Tu variants. His-tagged EF-Tu variants were mixed with either eukaryotic (eEF1B) or bacterial (EF-Ts) nucleotide exchange factors (with no His-tag) and incubated at room temperature for 4 hr and then passed-through Ni^{2+} column. To remove all non-specific interactions, the column was washed with a buffer prior to the elution. In the gel, top line represents EF-Tu variants and bottom line represents nucleotide exchange factors (EF-Ts, eEF1B). **A)** Knock-out EF-Ts from EF-Tu variants. **B)** Knock-in eEF1B into EF-Tu. **C)** Knock-out EF-Ts + Knock-in eEF1B into EF-Tu. For abbreviations, see table 2.2.

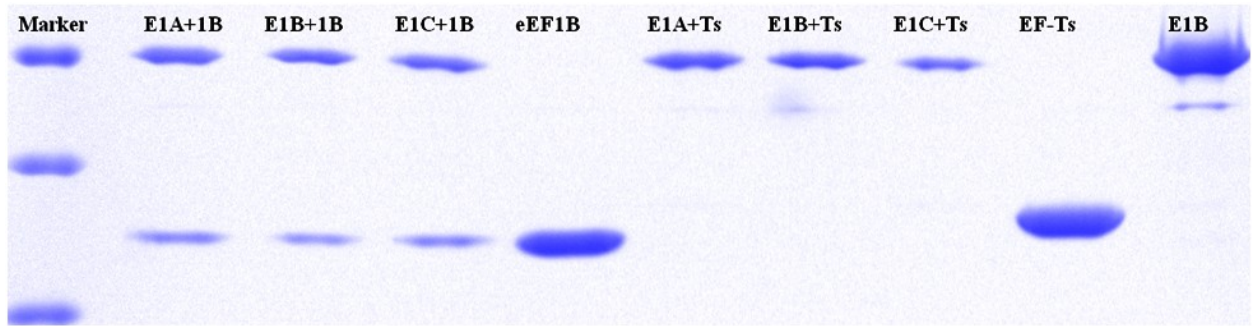
Binding ability of eEF1A Variants to EF-Ts or eEF1B

Binding ability of eEF1A variants were also determined by pull-down assay and the same parameters were used, as well. Knocking out of eEF1B from eEF1A results revealed that eEF1B was partially knocked-out and the rest of the protein remained bound to eEF1A, but the intensity of the bands were low while compare to the wild-type interaction of eEF1A:eEF1B. Conversely, no interaction of eEF1A:EF-Ts was observed and this also validate that there is no background interaction between these two proteins (Figure 2.18A).

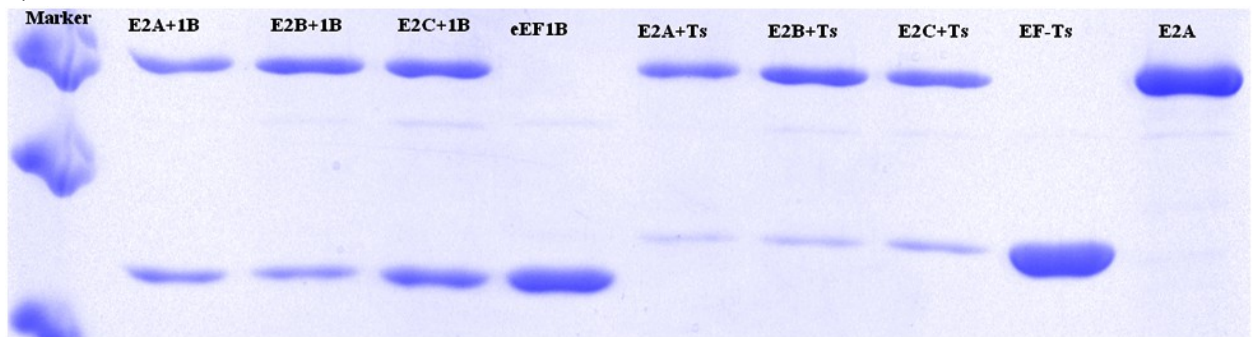
When we knocked-in EF-Ts into eEF1A, we observed some bands that indicate the ability of EF-Ts to bind eEF1A (Figure 2.18B).

The knocking out eEF1B plus knocking in EF-Ts results demonstrated that EF-Ts showed great dependency to eEF1A. The amount of EF-Ts that bound eEF1A had greatly increase. Especially, EF-Ts bound a lot to E3C (Type 1 divergence with $PP \geq 0.80$ + Type 2 divergence).

A)



B)



C)

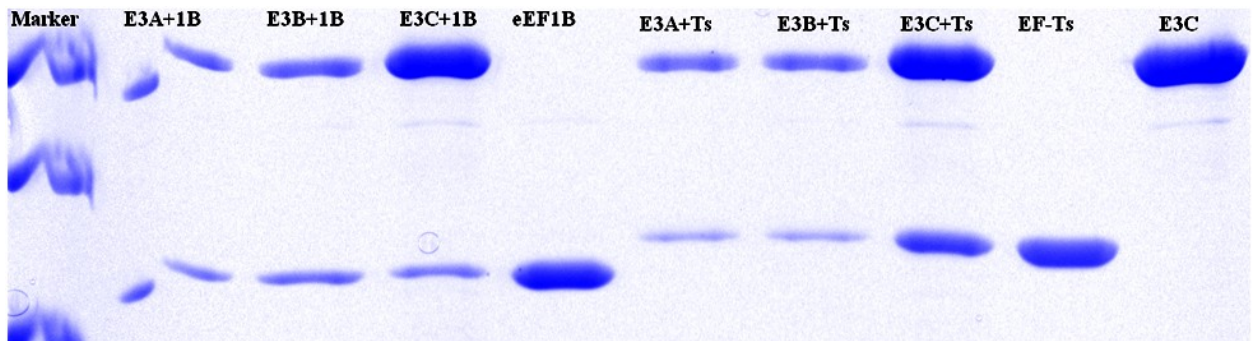


Figure 2.18. SDS-PAGE gel analysis of binding ability of nucleotide exchange factors to eEF-1A variants. His-tagged eEF1A variants were mixed with either eukaryotic (eEF1B) or bacterial (EF-Ts) nucleotide exchange factors (with no His-tag) and incubated at room temperature for 4 hr, and then passed-through Ni^{2+} column. Non-specific interactions were washed off with a buffer prior to the elution. In the gel, top line represents eEF1A variants and bottom line represents nucleotide exchange factors (EF-Ts, eEF1B). **A)** Knock-out eEF1B from eEF1A variants. **B)** Knock-in EF-Ts into eEF1A. **C)** Knock-out eEF1B + Knock-in EF-Ts into eEF1A. Abbreviations refer the parameters in Table 2.2.

2.4. CONCLUSION

We have experimentally validated site-specific rate shifts (type-I, type-II functional divergence) in the EF protein family. We used the power of statistical analysis and synthetic biology to engineer elongation factors to bind the non-cognate exchange factor in the other domain of life and vice-versa for bacteria and eukaryotes.

These findings may enable us to make functional inferences and explanations about minor functional diversification among EF-Tu and eEF1A because conservative in one site of a homologous proteins may acquire a new function, or becoming more variable may lead to a loss in function. In addition, our results confirm that heterotachy is an important pattern to generate functional diversity during the evolution of protein families.

Although the structures of eEF1A and EF-Tu are slightly different, they share similar mechanisms in their interactions with their cognate nucleotide exchange factors eEF1B and EF-Ts, which are not themselves homologous. When we look at the binding ability of nucleotide exchange factors to the variant, eEF1B did not display large variation in terms of binding EF-Tu. This could be because the sites where eEF1B interacts with eEF1A are conserved in both domains of life and eEF1B has some background interactions with wild-type EF-Tu. Conversely, EF-Ts showed large variability in its ability to bind to eEF1A variants.

For future directions, crystallization of some of the interactions between variants and nucleotide exchange factors would be valuable. This would provide us with a unique opportunity to further dissect molecular interactions that enable functional divergence.

In total, we have shown how an evolutionary synthetic biology can both generate biomolecules potentially useful for biomedicine and exploit evolutionary models to guide the

synthesis of molecules that allow us to better understand functional divergence among related sequences.

2.5. APPENDIX

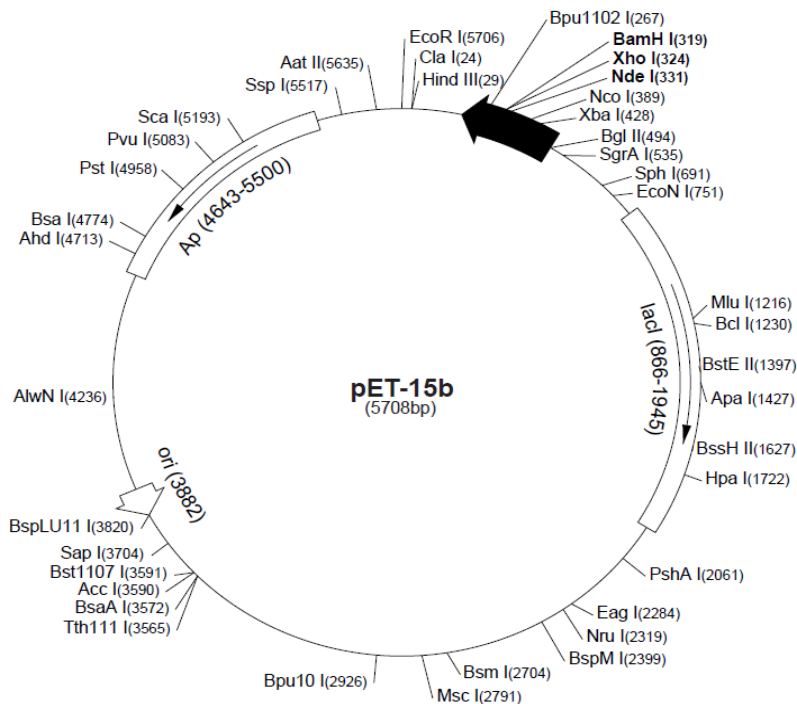
pET-15b Vector

TB045 12/98

The pET-15b vector (Cat. No. 69661-3) carries an N-terminal His•Tag® sequence followed by a thrombin site and three cloning sites. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below.

pET-15b sequence landmarks

T7 promoter	453-469
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites (<i>Nde</i> I - <i>Bam</i> H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
<i>bla</i> coding sequence	4643-5500



T7 promoter primer #69348-3
 T7 promoter
 lac operator
 Xba I
 rbs
 Bgl II
 AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGACGGGATAACAATCCCTCTAGAAATAATTTGTTTAACCTTAAGAGGAGA
 Nco I
 His•Tag
 Nde I Xho I BamH I
 TATACCATGGGACGAGCCATCATCATCATCACAGCAGCGGCTGGTGCCTGCGCGGAGCCATATGCTCGAGGATCCGGCTGCTAACAAAGCCCGA
 MetGlySerSerHisHisHisHisHisSerSerGlyLeuValProArgGlySerHisMetLeuGluAspProAlaAlaAsnLysAlaArg
 Bpu1102 I
 thrombin
 T7 terminator
 AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGACCAATAACTAGCATAACCCCTTGGGGCTCTAAACGGGCTTTGAGGGGTTTTTTG
 LysGluAlaGluLeuAlaAlaAlaThrAlaGluGlnEnd
 T7 terminator primer #69337-3

pET-15b cloning/expression region

pET-21a-d(+) Vectors

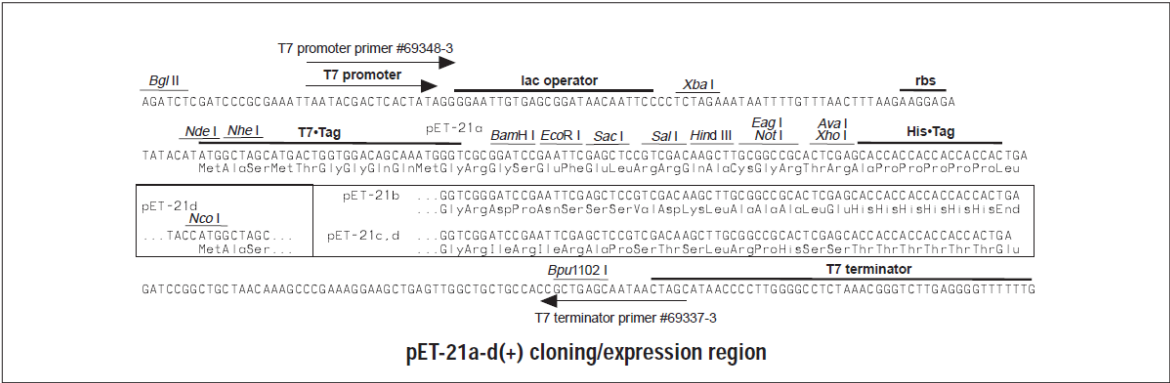
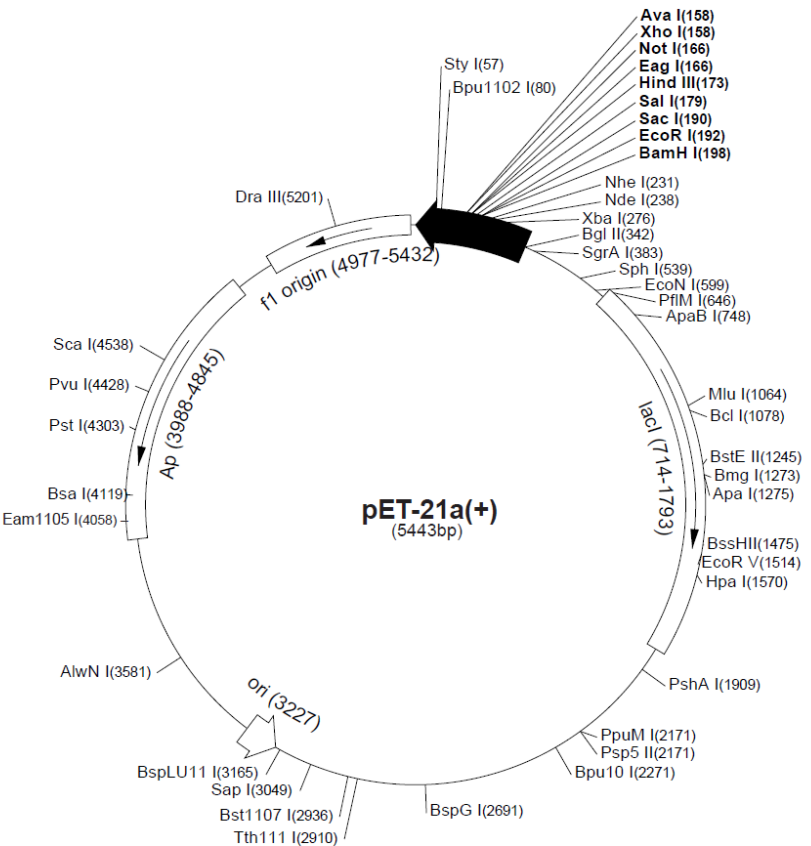
TB036 12/98

	Cat. No.
pET-21a DNA	69740-3
pET-21b DNA	69741-3
pET-21c DNA	69742-3
pET-21d DNA	69743-3

The pET-21a-d(+) vectors carry an N-terminal T7•Tag® sequence plus an optional C-terminal His•Tag® sequence. These vectors differ from pET-24a-d(+) only by their selectable marker (ampicillin vs. kanamycin resistance). Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).


pET-21a(+) sequence landmarks	
T7 promoter	311-327
T7 transcription start	310
T7•Tag coding sequence	207-239
Multiple cloning sites	
(BamH I - Xho I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	714-1793
pBR322 origin	3227
bla coding sequence	3988-4845
f1 origin	4977-5432

The maps for pET-21b(+), pET-21c(+) and pET-21d(+) are the same as pET-21a(+) (shown) with the following exceptions: pET-21b(+) is a 5442bp plasmid; subtract 1bp from each site beyond BamH I at 198. pET-21c(+) is a 5441bp plasmid; subtract 2bp from each site beyond BamH I at 198. pET-21d(+) is a 5440bp plasmid; the BamH I site is in the same reading frame as in pET-21c(+). An Nco I site is substituted for the Nde I site with a net 1bp deletion at position 238 of pET-21c(+). As a result, Nco I cuts pET21d(+) at 234, and Nhe I cuts at 229. For the rest of the sites, subtract 3bp from each site beyond position 239 in pET-21a(+). Nde I does not cut pET-21d(+). Note also that Sty I is not unique in pET-21d(+).



List of Sequences for Knock-out variant design

Sites with greatest statistical support from Diverge Analysis that are within 5 Angstroms of respective nucleotide exchange factors

Bacteria (1EFU)		Eukaryotes (Yeast: 1F60 & 2B7C)	Key
Align #	Type I (Gu 99 & 01) & Type II	Type I (Gu99 & 01)	
10	16THR	V	DIVERGE Results Color PP  100-90  89-80
23	29ALA	G	
26	32THR	D	Type II Diverge  59.847  19.75
27	33THR	K	
31	37LYS	E	Other Residue 
36	42A	E	
37	43A	F	
38	44R	K	
39	45A	Y	
40	46F	A	
41	47D	W	
42	48Q	V	
43	49I	L	
45	51N	K	
47	53P	64LYS	
48	54E	A	
50	56K	67ARG	
51	57A	68GLU	
57	63N	74ASP	
58	T	75ILE	
60	66HIS	77LEU	
72	78HIS	89VAL	
73	79VAL	90ILE	
89	A	106THR	
90	A	107SER	
102	108THR	G	
103	109ASP	V	
105	111PRO	K	
106	112MET	D	
133	139MET	S	
138	144GLU	S	
139	145LEU	R	
142	148LEU	E	
143	149VAL	I	

171	179GLU	K
172	180GLY	T
204	E	249GLN
207	F	252TYR
208	S	253LYS
212	R	257ILE
217	T	262VAL
248	F	293HIS
263	V	308PHE
264	L	309ASN
310	323PHE	L
311	K	360ASP
335	348GLU	K
336	349MET	F
337	350VAL	L
367	G	428ARG
368	R	429GLN

List of Sequences for Knock-in variant design

In Pymol aligned the 1EFU & 1F60 structures, identified those sites that were within 5 Angstroms of the opposite nucleotide exchange factor, and selected those with the highest statistical support

Bacteria (1EFU)		Eukaryotes (Yeast: 1F60)	Key	
Align #	Type I (Gu 99 & 01) & Type II	Type I (Gu99 &01) & Type II	DIVERGE Results	
16	H	18SER	Color	PP
21	L	23THR		100-90
37	A	54PHE		89-80
38	R	55LYS		
39	A	56TYR		
40	F	57ALA		
41	D	58TRP		
43	I	60LEU		
47	P	64LYS		
51	A	68GLU		
58	64THR	I		

Type II Diverge



Other Residue



60	66HIS	L
62	68GLU	K
72	78HIS	V
73	79VAL	I
89	95ALA	T
98	V	115ILE
103	D	120VAL
105	P	129LYS
106	M	130ASP
138	E	163SER
139	L	164ARG
141	E	166GLN
142	L	167GLU
145	M	170LYS
150	L	175PHE
153	Q	178LYS
154	Y	179VAL
166	A	193GLY
167	L	194TRP
169	A	196GLY
204	215GLU	Q
248	261PHE	H
250	263LYS	E
259	272GLU	D
299	312SER	346ASN
301	314ASP	348PRO
302	315GLU	349GLY
303	316GLY	350GLN
304	317GLY	I
339	P	394SER
366	G	427MET
367	380GLY	R
368	381ARG	Q
371	394GLY	A

2.6. BIBLIOGRAPHY

Andersen, G. R., L. Valente, et al. (2001). "Crystal structures of nucleotide exchange intermediates in the eEF1A-eEF1B α complex." *Nat Struct Biol* **8**(6): 531-534.

Baldauf, S. L., J. D. Palmer, et al. (1996). "The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny." *Proc Natl Acad Sci U S A* **93**(15): 7749-7754.

Borradaile, N. M., K. K. Buhman, et al. (2006). "A critical role for eukaryotic elongation factor 1A-1 in lipotoxic cell death." *Mol Biol Cell* **17**(2): 770-778.

Cole, M. F. and E. A. Gaucher (2011). "Exploiting models of molecular evolution to efficiently direct protein engineering." *J Mol Evol* **72**(2): 193-203.

Fitch, W. M. and E. Markowitz (1970). "An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution." *Biochem Genet* **4**(5): 579-593.

Gaucher E.A., Gu X., Miyamoto M.M., Benner S.A. Predicting functional divergence in protein evolution by site-specific rate shifts (2002) *Trends in Biochemical Sciences*, 27 (6), pp. 315-321.

Gaucher, E. A., M. M. Miyamoto, et al. (2001). "Function-structure analysis of proteins using covarion-based evolutionary approaches: Elongation factors." *Proc Natl Acad Sci U S A* **98**(2): 548-552.

Gaucher, E. A., U. K. Das, et al. (2002). "The crystal structure of eEF1A refines the functional predictions of an evolutionary analysis of rate changes among elongation factors." *Mol Biol Evol* **19**(4): 569-573.

Goldman, N. and S. Whelan (2000). "Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics." *Mol Biol Evol* **17**(6): 975-978.

Gonen, H., C. E. Smith, et al. (1994). "Protein synthesis elongation factor EF-1 alpha is essential for ubiquitin-dependent degradation of certain N alpha-acetylated proteins and may be substituted for by the bacterial elongation factor EF-Tu." *Proc Natl Acad Sci U S A* **91**(16): 7648-7652.

Gross, S. R. and T. G. Kinzy (2005). "Translation elongation factor 1A is essential for regulation of the actin cytoskeleton and cell morphology." *Nature Structural & Molecular Biology* **12**(9): 772-778.

Gu, X. (2006). "A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences." *Mol Biol Evol* **23**(10): 1937-1945.

Gu, X. (2003). "Functional Divergence in Protein (Family) Sequence Evolution." *Genetica* **118**(2): 133-141.

Gu, Xun, Maximum-Likelihood Approach for Gene Family Evolution Under Functional Divergence, *Mol. Biol. Evol.* **18**(4):453–464. 2001.

Gu, X. and Vander Velden, K. (2002) DIVERGE: phylogeny-based analysis for functional–structural divergence of a protein family. *Bioinformatics* **18**, 500–501

Hervé Joël Defeu Soufoa, Christian Reimolda, Uwe Linneb, Tobias Knusta, Johannes Geschera, and Peter L. Graumann (2010). Bacterial translation elongation factor EF-Tu interacts and colocalizes with actin-like MreB protein, *PNAS*. Vol107.

Hotokezaka, Y., U. Tobben, et al. (2002). "Interaction of the eukaryotic elongation factor 1A with newly synthesized polypeptides." *J Biol Chem* **277**(21): 18545-18551.

Kjeldgaard, M., J. Nyborg, et al. (1996). "The GTP binding motif: variations on a theme." *FASEB J* 10(12): 1347-1368.

Lodish, H. et al. (2008) Molecular Cell Biology, 6th edition. W.H. Freeman and Company, New York.

Lopez, P., D. Casane, et al. (2002). "Heterotachy, an important process of protein evolution." *Mol Biol Evol* **19**(1): 1-7.

Lopez, P., P. Forterre, et al. (1999). "The root of the tree of life in the light of the covarion model." *J Mol Evol* **49**(4): 496-508.

Miyamoto, M. M. and W. M. Fitch (1995). "Testing the covarion hypothesis of molecular evolution." *Mol Biol Evol* **12**(3): 503-513.

Philippe, H., P. Lopez, et al. (2000). "Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions." *Proc Biol Sci* **267**(1449): 1213-1221.

Philippe, H. and P. Lopez (2001). "On the conservation of protein sequences in evolution." *Trends Biochem Sci* **26**(7): 414-416.

Revell, L. J., L. J. Harmon, et al. (2008). "Phylogenetic signal, evolutionary process, and rate." *Syst Biol* **57**(4): 591-601.

Wang, Q. F., S. Prabhakar, et al. (2006). "Primate-specific evolution of an LDLR enhancer." *Genome Biol* **7**(8): R68.

Yang Z. (1996b). "Maximum-likelihood models for combined analyses of multiple sequence data." *J Mol Evol.* 42:587–596