

Reengineering Coding Specificities in Glutamyl-
and Glutamyl-tRNA Synthetases

Thesis Dissertation

Doctorate of Philosophy – Biochemistry and Molecular Biology

Andrew Hadd

Advisor: John Perona, Ph.D.

Oregon Health and Science University

Department of Biochemistry and Molecular Biology

Portland, Oregon

December 2014

Department of Biochemistry and Molecular Biology

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Doctoral thesis of

Andrew Hadd

has been approved

John Perona, Ph.D. - advisor

David Farrens, Ph.D.

Dirk Iwata-Reuyl, Ph.D.

Matt Thayer, Ph.D.

Abstract

Translating nucleic acid based information into a protein product is at the heart of the genetic code. Aminoacyl-tRNA synthetases (aaRSs) function is the basis of decoding the genetic code by ligating the cognate tRNA and amino acid together. Given this role, aaRSs maintain an enormous level of specificity for their amino acid and tRNA substrates. This thesis sought to uncover the mechanisms of substrate specificity in aaRSs using glutamyl- and glutamyl-tRNA synthetases (GlnRS and GluRS, respectively) as model systems. The work addresses the macromolecular coevolution in this subfamily, revealing the nucleic acid recognition signatures within the tRNA utilized by eukaryotic GlnRS and GluRS. Furthermore a rational engineering method discovered that the aforementioned tRNA signatures also assist the corresponding GlxRS in amino acid recognition, a novel concept in genetic code translation. These generalized principles and engineering strategy can be applied to further our understandings in protein-nucleic acid readout mechanisms as well as achieving goals in genetic code expansion for synthetic biology efforts.

Acknowledgments

Family, friends and lab mates :

These people and this experience have fundamentally shaped my life.

Publication List

Hadd, A. & Perona, J.J. – *Recoding aminoacyl-tRNA synthetases for synthetic biology by rational protein-RNA engineering* (2014) ACS Chemical Biology *accepted* – Data in Chapter II

Hadd, A. & Perona, J.J. – *Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl-tRNA synthetases* (2014) Journal of Molecular Biology 426(21):3619-3633– Data in Chapter I

Review Article - Perona, J.J. & Hadd, A. – *Structural diversity and protein engineering of the aminoacyl-tRNA synthetases* (2012) Biochemistry 51(44):8705-8729

Grant, T.D., Snell, E.H., Luft, JR., Quarterly, E., Corretore, S., Wolfley, J.R., Snell, M.E., Hadd, A., Perona, J.J., Phizicky, E.M., and Grayhack E.J. – *Structural conservation of an ancient tRNA sensor in eukaryotic glutaminyl-tRNA synthetases* (2012) Nucleic Acid Research 40(8):2723-2731 – Data in Chapter I

Rodriguez-Hernandez, A., Bhaskaran, H., Hadd, A., and Perona, J.J. – *Synthesis of Glu-tRNA(GLN) by engineered and natural aminoacyl-tRNA synthetases* (2010) Biochemistry 10:49(31):6727-6736 – Data in Chapter II

Tables of Contents

Introduction.....	1
Chapter I – <i>Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl-tRNA synthetases</i>	31
Chapter II – <i>Recoding aminoacyl-tRNA synthetases for synthetic biology by rational protein-RNA engineering</i>	56
Chapter III – Crystallography of <i>Sacchromyces cerevisiae</i> Arc1p-GluRS-tRNA ^{Glu} complex.....	71
Chapter IV – Concluding remarks.....	78
Methods.....	85
References.....	99

---- Figures List ----

Figure 1 – Codon table – 2

Figure 2 – Secondary and tertiary tRNA structure – 2

Figure 3 – Aminoacylation mechanism – 6

Figure 4 – aaRS class defining structural elements – 7 & 8

Figure 5 – Cys-tRNA^{Cys} biosynthesis – 10

Figure 6 – Multi aminoacyl-tRNA synthetases complex components – 15

Figure 7 – Human GlnRS regulating Ask1 activity – 16

Figure 8 – Glu-tRNA^{Glu} and Gln-tRNA^{Gln} biosynthesis – 20

Figure 9 – Phylogenetic tree of GluRS subfamily – 21

Figure 10 – Acceptor arm recognition in *E. coli* GlnRS – 24

Figure 11 – Representative aminoacylation kinetics plots – 39

Figure 12 – Cloverleaf depiction of eukaryotic tRNA^{Glx} identity elements – 43

Figure 13 – Structural overlay of GlnRS and GluRSND – 44

Figure 14 – Unrooted phylogenetic tree of all tRNA^{Glx} – 52

Figure 15 – *E. coli* GlnRS structure – 59

Figure 16 – Amino acid binding pocket of GluRS and GlnRS sequence alignment – 60

Figure 17 – Reengineering tRNA identity elements for hybrid system – 63

Figure 18 – Free energy analysis of reengineered GlnRS-tRNA complex – 65

Figure 19 – Aminoacyl-tRNA synthetase engineering strategy – 68

Figure 20 – Interacting domains of *S. cerevisiae* GluRS-Arc1p-MetRS complex – 72

---- Table List ----

Table 1 – *S. cerevisiae* GlnRS and GluRS aminoacylation kinetics – 38

Table 2 – Probing functional importance of tRNA^{Glx} identity elements on charging – 42

Table 3 – Steady state kinetics effects on tRNA identity swap experiments – 46

Table 4 – Steady state kinetics of reengineered GlnRS and tRNA – 62

Introduction

Translation and the genetic code

The central dogma of biology is the cornerstone for biological entities to convert a nucleic acid based message into the genetically coded protein product. Contemporary life utilizes deoxyribonucleic acid (DNA) as the storage molecule to be converted into the intermediary ribonucleic acid (RNA), which is translated into a protein. The basis for life to propagate and evolve rests upon a cell's ability to maintain and carryout this process. However, for protein synthesis to occur, the cell has to translate a coded message of one molecule type into the product of another; thus Nature evolved the genetic code, the conversion of nucleic acid message into an amino acid product.

Protein synthesis occurs on the ribosome, a complex RNA-based molecular machine. It is built with two major molecular subunits and synthesizes protein converted from the intermediary messenger RNA (mRNA) utilizing the genetic code. The ribosome "reads" the mRNA nucleotides in triplets called codons. Thus every three nucleotides represent an amino acid. In order to link to the molecules together, the mRNA and amino acid, a bridging molecule is needed and accomplished by a transfer RNA (tRNA) molecule. A tRNA is small RNA molecule (~76nts), which gets ligated to a corresponding amino acid. Within the tRNA, there are complimentary nucleotide triplets that will Watson-Crick base pair to the mRNA codon, called an anticodon. This base pairing is how the tRNA acts to bridge the codon with its coded amino acid and each

anticodon corresponds to one of the twenty amino acids. Thus the genetic code is a set of triplet nucleotide sequences that represent a particular amino acid. The genetic code is broken down into sixty-four codons that are divided into corresponding amino acids. The code is degenerate meaning multiple codons can represent the same amino acid, however, is not evenly distributed. Some amino acids can have up to six codons and as few as one (Figure 1). Within the code exists a single start codon, AUG representing methionine and also three separate stop codons to signal translation termination.

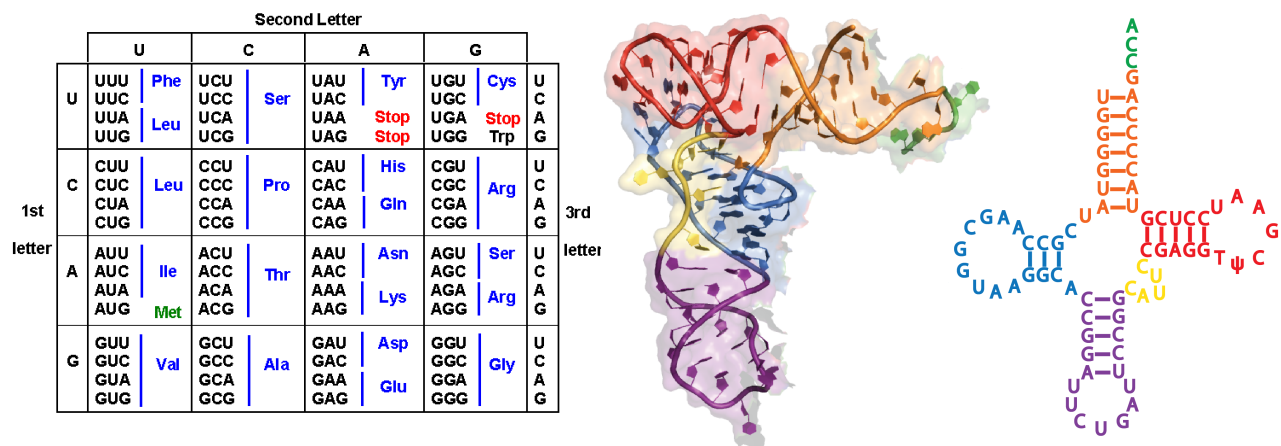


Figure 1 (left) – Genetic code codon table showing distribution of nucleic acid sequence representing the particular amino acid (openwetware). Figure 2 (right) - Depiction of the secondary (right) and tertiary (left) tRNA structure. Each color represents the different domains: acceptor arm (orange), D arm (blue), anticodon arm (purple), variable loop (yellow), TψC arm. The conserved modified ψ55 in the T-arm is labeled as such. Each tRNA ends in CCA (green). The tRNA is a representation of *E. coli* tRNA^{Gln}.

The tRNA molecule may be a relatively small molecule, between 73 and 103 nucleotides long, but possess extraordinary capabilities. Of note, the conventional numbering system ends with 76 nucleotides and will be used throughout this work. tRNAs have a generalized secondary structure in a cloverleaf configuration and adopts a native L-shaped tertiary structure upon folding (Figure 2). Each tRNA, regardless of which amino acid it represents, is indeed generalized, as they are all substrates for the

ribosome and elongation factor during protein synthesis. The four tRNA arms, (acceptor, D-, TΨC- and anticodon) contain both elements that give the molecule its general features but also sequence signatures making them unique. One general feature includes all tRNAs end in CCA, where the 3'hydroxyl of A76 attaches to the amino acid. In the natively L-shaped folded tRNA molecule, the horizontal arm has twelve stacking layers and the vertical arm has fourteen stacking layers containing both base pairs and triples. The hinge is created through interactions of the D-arm and TΨC-arm through interactions with conserved nucleotide elements (Figure 2). These tRNA elements include G18 and G19 in the D-loop, which interact with T54 Ψ55 C56 in the T-loop (Figure 2). Of note, Ψ55 is a universally conserved post-transcriptional modification introduced by tRNA pseudouridine synthase from genetic encoded U55 [1, 2]. Examples of these commonly conserved elements allow for the tRNA to maintain an overall generalized shape allowing the molecule to be amenable to the universal translation system or the common set of tRNA modifying enzymes that synthesize the universal post-transcriptional modifications.

tRNA molecules are some of the most post-transcriptionally modified RNA molecules within the cell [2]. Granted this can be justified given the central role the tRNA plays in maintaining the genetic code. However an abundance of energy is dedicated to these modifications, many requiring multiple enzymes to generate the final product. As already mentioned, some modifications are found in all tRNAs while others are tRNA specific [2, 3]. Lastly, some tRNA modifications may be kingdom specific, as some are only found in bacteria, archaea or eukaryotes. Nonetheless, modifications

further assist in generating the precise level of diversity and similarity for the tRNA to perform its function.

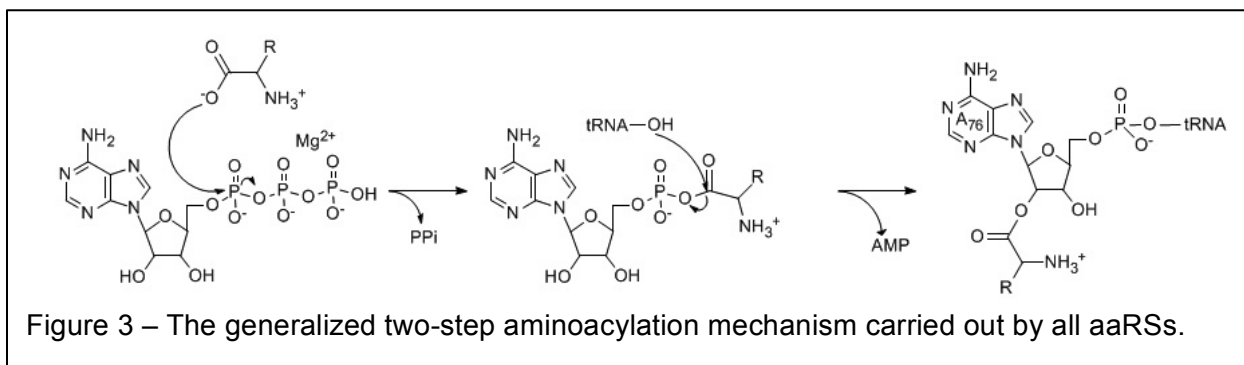
Anyone working with tRNA must consider post-transcriptional modifications for their experiments given their pivotal role in biology. One example is a hypermodified base found in all tRNA^{Glu}, tRNA^{Gln} and tRNA^{Lys}, all three of which are tRNA isoacceptors with a uracil at wobble position 34 in their anticodons [2, 3]. Furthermore this base is modified in all organisms, but is slightly different amongst the three phylogenetic domains. The first modification is a sulfur atom that replaces the oxygen at the 2-position. The 5-position is either modified to methylaminomethyl or carboxymethylaminomethyl. This modification plays two central roles in decoding; it is important for recognition towards its specific aminoacyl-tRNA synthetase and also important for interactions on the ribosome assisting in the proper orientation and interactions with the cognate codon [4-7].

However, since the fidelity of the genetic code is fundamentally dependent upon the correct coupling between the amino acid and tRNA, each tRNA molecule must also possess nucleotides to distinguish them. An obvious example is the anticodon sequence, since it pairs with the mRNA codon representing the required amino acid. Other distinguishing elements are typically found within the acceptor arm. These unique signatures ensure the proper specificity between the RNA molecule and any potential binding partners (e.g. aminoacyl-tRNA synthetases or *specific* tRNA modifying enzymes). Furthermore, tRNA shape also plays a substantial role in generating the required level of specificity. For example, the specific elements in the acceptor arm in order to interact properly with a specific binding protein, the orientation of those

nucleotides must be oriented in a highly specific region in three-dimensional space. Therefore, elements within the rest of the tRNA molecule co-varied with those signatures that allows for this specific orientation. These elements may be found adjacent to the signatures or can be found angstroms away. An example of this will be discussed in Chapter II but briefly, RNA elements within the D-loop assist in orienting nucleotide signatures in the acceptor arm of tRNA^{Glu} that are required to interact with a specific protein. These concepts will also be discussed in detail throughout this work - how distinguishing elements is the heart of the genetic code, as those RNA signatures are what represent (*or code*) for a specific amino acid.

Aminoacyl-tRNA synthetases

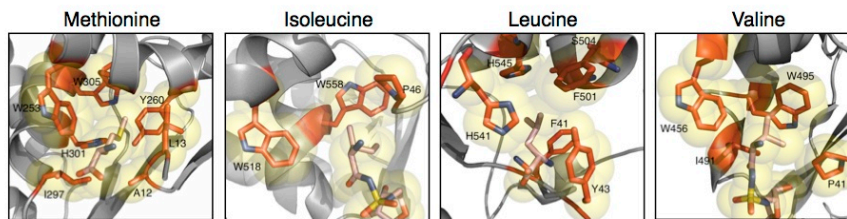
Although since the epicenter for protein synthesis occurs at the ribosome, a fundamental question can be asked, does *translation* truly occur there? By definition, translating the genetic code is the conversion of a nucleic acid message into coded amino acid product. Taking this definition, as already discussed, the cornerstone molecule is the tRNA as its genetically encoded sequence represents one of the twenty amino acids. Therefore the chemical process, which links the two molecules together, ought to be considered a true step of translation. Aminoacyl-tRNA synthetases (aaRSs) are the enzymes that carryout this process. A specific aaRS has therefore evolved to specifically recognize a specific tRNA and corresponding amino acid. This means, conventionally: twenty amino acids, twenty tRNA isoacceptors, and twenty aaRSs.



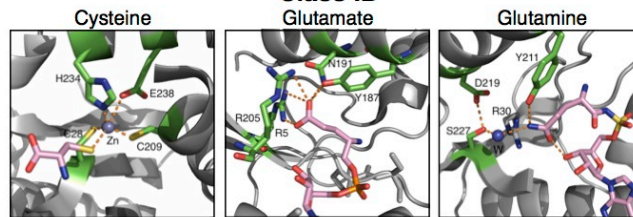
AaRSs ligate the amino acid and tRNA together through a two-step process called aminoacylation. First the amino acid is activated by an adenylation reaction using ATP before being transferred to the 3' end of the tRNA (Figure 3) [1]. While all aaRSs utilize the same aminoacylation mechanism, they are broken into two sub-classes, denoted class I and II. A detailed structural phylogenetic analysis was carried out by O'Donoghue et. al. in 2003, in which each amino acid specifying aaRS utilizing structural homology [8]. While sequence based alignments had been used previously for aaRSs, utilizing structural alignments to discern phylogeny and generate classifications is more accurate as structures are overall more conserved than sequence [8]. This phylogenetic study was made possible now that a representative crystal structure within the diverging classes of each aaRS spanning all domains of life and various species have been solved. This work revealed previously the two known main classes (class I and class II). However this detailed study now allows a true characterization for each sub-class for which each aaRS falls into. These findings revealed homologous structural elements within the catalytic sites generating each sub-class (Figure 4).

Class I

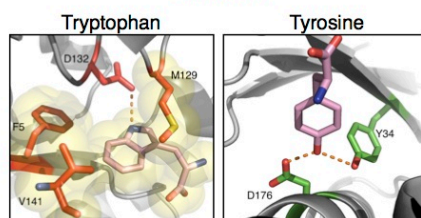
Class IA



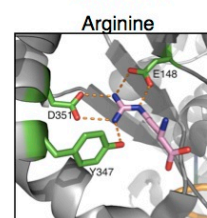
Class IB



Class IC

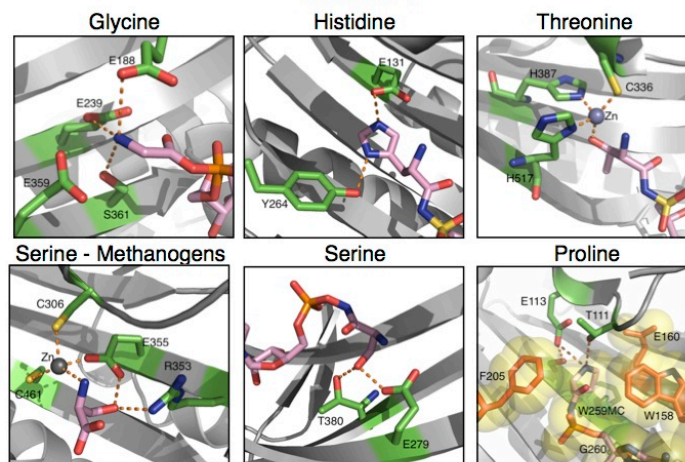


Class ID

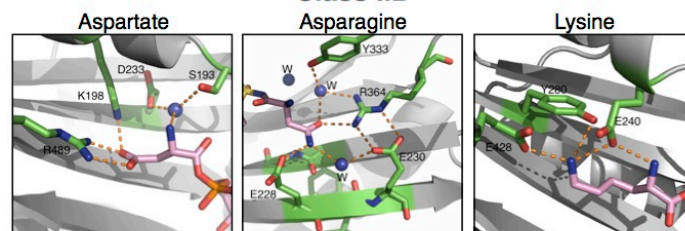


Class II

Class IIA



Class IIB



Class IIC

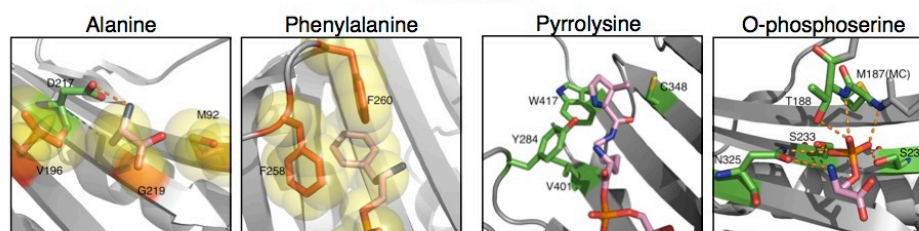
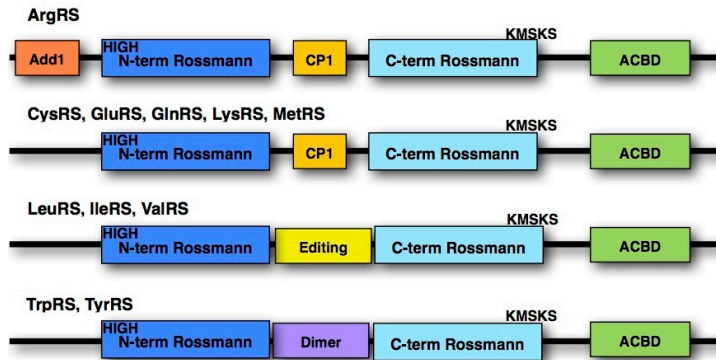
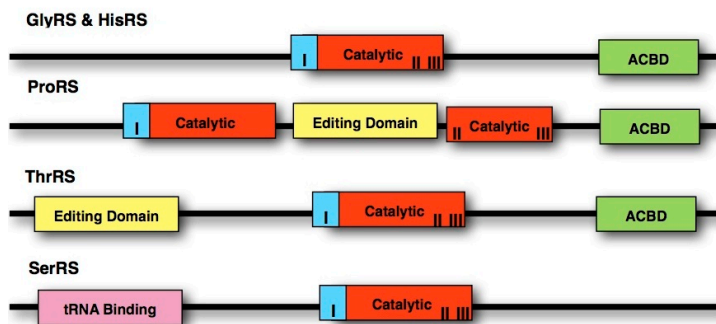


Figure 4 – Representative examples of each aaRS, categorized into sub-classes relative to amino acid specificities. AaRSs are shown in grey with protein residues that generate substrate interactions (H-bonding, green and van der Waals, red). Bound substrates are all shown in pink and are either free amino acid (Met, Leu, Cys, Trp, Tyr, Arg, Ser, Lys, Ala and Sep) or aminoacyl-adenylate intermediate analogs (Gln, Glu, Ile, Val, His, Thr, Ser, Gly, Pro, Asp, Phe and Pyl) [1].

Class I aaRS



Class IIA aaRS



Class IIB aaRS



Class IIC aaRS

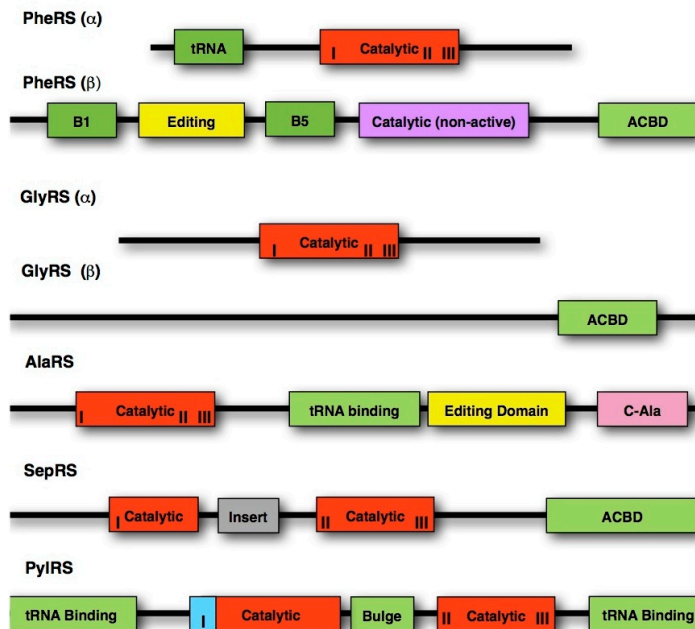


Figure 4 cont –

Primary sequence organization of each class of aaRS. Class I contain an N-terminal Rossmann fold, split by a CP1 domain. CP1 assist in orienting the tRNA into the active site. Furthermore, in Leu, Ile and Val CP1 contains post-transfer editing functions. In Tyr and Trp, CP1 generates the dimer interface, as they are the only two non-monomeric class I aaRSs. The ATP binding site is generated by conserved sequences motif, HIGH and KMSKS. Each class I aaRS also utilizes a C-terminal anticodon binding domain. ArgRS has an N-terminal domain (ADD1) that assists in tRNA interactions, creating an interface along tRNA^{Arg} core.

Class II is mostly found as either dimers or tetramers. As opposed to a Rossmann fold active site, class II have an active site with seven parallel β -sheets flanked by two α -helices, denoted as catalytic domain (red). The catalytic motifs can be broken into thirds, noted by I, II, and III. In class IIa, each have a C-terminal anticodon binding domain, except for SerRS, which is replaced with an extra N-terminal tRNA binding domain. Editing domain insertion in ProRS is found within the catalytic domain, while ThrRS has an N-terminal editing site. Class IIb each have N-terminal anticodon binding domain and C-terminal active site and are all dimers. Finally the divergent class IIc. Tetrameric PheRS and GlyRS is a dimer of heterodimers ($\alpha\beta$). PheRS β -chain contains post-transfer editing function. AlaRS is interface, which is created by the C-terminal Ala specific domain. SepRS is a homotetramer (α_4). PylRS is a homodimer and does not interact with the tRNA anticodon but instead has two-tRNA binding domains. [1]

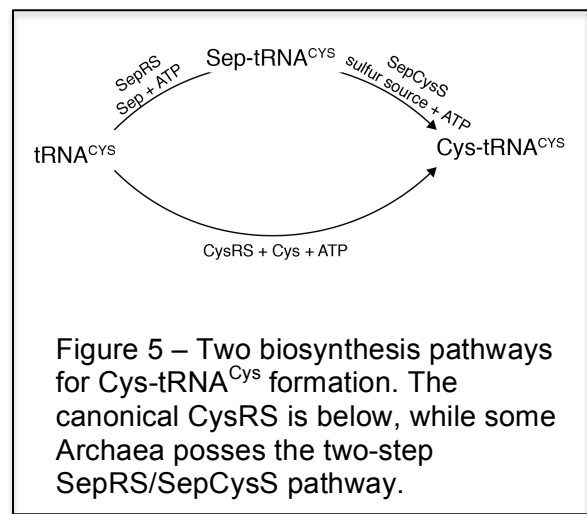
Class I are typically monomeric enzymes, each with N-terminal Rossmann fold active site and a C-terminal RNA binding domain for anticodon recognition. The peptide chain forming the Rossmann fold is split in half by a connective peptide domain (CP1) assisting to orient the tRNA into the active site [1]. ATP binding is found with sequence signatures, HIGH and KMSKS. The HIGH is found in the first half of the Rossmann fold and KMSKS is in a loop at the C-terminus of the active site. Amino acid binding sites are obviously distributed throughout the active site (Figure 4).

Class I aaRSs can be further broken down into four separate groups: Ia, Ib, Ic, and Id (Figure 4) [1]. These subclasses are divided based on commonalities in their active site structure (Figure 4); for example, TyrRS and TrpRS are class Ic, both possessing bulky ring structure amino acid substrates. Class Ia LeuRS, IleRS, ValRS are unique due to an enlarged CP1 domain that contains catalytic activity possessing editing functions. These editing functions allow these three aaRSs to proofread the aminoacylated tRNA product ensuring the correct amino acid was ligated. Thus these three aaRSs with aliphatic amino acid specificities have a propensity to aminoacylate their respective tRNA with an incorrect amino acid, necessitating this editing activity. The only exception for the monomeric quaternary structure are within the class Ic is tyrosyl-tRNA synthetase (TyrRS) and tryptophanyl-tRNA synthetase (TrpRS), which are each dimers. As with the three editing class Ia, the TyrRS and TrpRS dimer interface resides within the enlarged CP1 domain.

Class II aaRSs are more distinct in quaternary structure organization and exist as dimers or tetramers. As opposed to their class I counterparts, class II have an active

site with seven parallel β -sheets flanked by two α -helices [1, 8]. Class II are further divided into three groups: IIa, IIb and IIc. As with class I sub-classifications, class II are classified based on structural conservation patterns within the active site (Figure 4). Class IIa are distinct due to a structurally conserved tRNA binding domain, albeit not conserved in relative placement within the gene (Figure 4). Class IIb are dimers with C-terminal tRNA binding domains, each specific for large polar amino acids. Class IIc bear little commonalities amongst each other and are either homodimers (α_2) or heterotetramers ($\alpha_2\beta_2$). Unlike class IIa and IIb which either have common tRNA binding or similar amino acid chemistries to aid classification, IIc are much more divergent. PheRS and AlaRS both are $\alpha_2\beta_2$ tetramers and have editing domains. The protein-protein interfaces and RNA binding domains are intertwined throughout the quaternary structure and highly complex and unlike any binding mode in IIa or IIb (Figure 4). Furthermore, class IIc breaks the rule that each aaRS is found as only one subtype; the exception is GlyRS. Little information is known about class IIc GlyRS other than the quaternary structure is $\alpha_2\beta_2$, but holds no common traits with the functionally homologous class IIa GlyRS.

Also in class IIc exists the two *non-standard* aaRSs, O-phosphoseryl-tRNA synthetase (SepRS) and pyrolysyl-tRNA synthetase (PylRS) [1]. The little similarities amongst class IIc highlights the plasticity in their structures. Both SepRS and PylRS are exclusively found in archaea and are related



to unique attributes within the host organism metabolism. With some archaea living in anaerobic environments with limiting nutrients, the translation systems must adapt. For example, sulfur and cysteine metabolism in some archaea is unique and thought to be of ancient origin [9]. Some organisms that lack canonical cysteine biosynthesis genes utilize multiple pathways to synthesize Cys-tRNA^{Cys} (Figure 5). SepRS ligates phosphorylated-serine (Sep) onto tRNA^{Cys}. The Sep-tRNA^{Cys} product is then a substrate for a second enzyme, SepCysS, to synthesize Cys-tRNA^{Cys} with a reduced sulfur atom. A by-product in this reaction is phosphate coming from the phosphorylated-serine.

LysRS is another system that can be found in multiple aaRS variants [1, 8]. Unlike GlyRS, where both variants are within class II, LysRS is found either as class I or II. Most organisms utilize class II variants that include all eukaryotes and most bacteria. Class I LysRS is found dispersed throughout the archaeal lineage, and furthermore, a small subset of archaeal organisms contain both a class I and II LysRS [1, 8]. The evolutionary and biological rationale for both LysRS classes is not understood.

The structural diversity gives aaRS various means to uniquely recognize their specific amino acid and tRNA substrates. This is accomplished through the recognition of specific stereochemical signatures within the substrates and defined as identity elements. Amino acid recognition strategies take advantage of the variation on the side chain chemistries and size (Figure 4). Amino acid selectivity also is accomplished by rejecting near cognate substrates. For example in PheRS, residues interact with phenylalanine substrate through a π -stacking interaction with two phenylalanine residues within the amino acid binding pocket [1]. Furthermore, the orientation of the

binding pocket is closed at the far end, which could create a steric clash with the para-hydroxyl of tyrosine, thus rejecting it.

However, aaRSs that have the potential to mis-aminoacylate their cognate tRNA with a near but non-cognate amino acid have evolved editing activities. These second catalytic activities can either occur within the same synthetic site or a second catalytic site. Editing activities in aaRSs can occur after either step in aminoacylation, termed pre-transfer or post-transfer, respectively. The combination of highly specific amino acid binding sites that are potentially supplemented with editing activities gives aaRS enough specificity to maintain an error rate of one misincorporation every 10^4 events [1].

A similar strategy for tRNA recognition can be made too. Cognate tRNA and aaRS have coevolved, resulting in a relatively tight binding constant ($K_d \sim 10^{-7}$). tRNA substrates are recognized in a class specific manner, along the minor or major groove for class I and II, respectively. Most of tRNA discrimination elements reside within the acceptor arm and anticodon and are recognized by various mechanisms. These can include direct readout of specific bases and indirect interactions from the phosphate backbone [1, 10]. Indirect readout has been found to contribute an enormous level of specificity. These interactions are dependent upon the shape or dynamics of the RNA molecule, which are derived by the nucleotide sequence. This is a means to maximize the amount of specificity from a relatively small RNA molecule that must maintain overall generalized features for its interactions with other translation functions.

Recognition and discrimination between the tRNA and aaRS relies on negative selection as well as positive [1, 3]. Positive selection is relatively straightforward, meaning interactions between each macromolecule adds to the overall binding affinity

or assists in driving catalysis. However, these can be complicated in that positive identity elements do not necessarily directly contact the protein and influence specificity through an allosteric pathway. For example in CysRS, the Levitt Pair (position 15-48) helps generate the shape of the tRNA core but no protein directly interacts with these nucleotide positions[10, 11]. The core shape however influences the orientation of the acceptor arm along the protein-RNA interface. Thus core nucleotides assists in allowing the readout of the acceptor arm and by definition both nucleotides signatures are considered positive identity elements. Anti-determinants are also found in tRNAs. These elements are important and help discriminate against near-cognate tRNA substrates. An example of this is found in tRNA^{Glu} against GlnRS, whose D-loops are relatively large compared to a near cognate tRNA^{Gln} [12]. Thus using a steric mechanism, the enlarged D-loop in tRNA^{Glu} selects against interactions with GlnRS.

As the genetic code evolved, these positive and negative determinants generate the foundation that maintains fidelity. These determinants coevolved together creating a dynamic scenario and are tuned to create the precise level of specificity and efficient rate of protein synthesis. The balanced amalgamation of positive and negative determinants together is what creates a cognate aaRS-tRNA pair. At a systems level, the coevolution of each identity set also ensures the rejection of non-cognate aaRSs by creating unfavorable interactions. These interactions have a clear positive selection related to correctly translated proteins carrying out their *in vivo* roles. Taken together a fundamental aspect translating the genetic code is the readout of the tRNA molecule. When classically defined, the genetic code is the 64-codon table representing a particular amino acid. However one can also account for the other 73 nucleotides within

the molecule as contributions towards coding. De Duvé noted that a second non-degenerate genetic code might lie within the tRNA [13]. Unlike the anticodon, which is not species dependent, other unique tRNA signatures by definition can still represent an amino acid. In this case, the genetic code is a nucleotide signature that gives the tRNA identity amongst the pool of other tRNAs. This unique signature does indeed code for the amino acid based upon the tRNAs functional role, linking codon to amino acid.

Interestingly, while the code is translated accurately, miscoding events can and do occur, continuing to propagate an evolving system. At times, miscoding events are advantageous. For example as seen in *Saccharomyces cerevisiae*, MetRS mis-methioninylates tRNAs other than tRNA^{Met} upon cellular oxidative stress [14]. The increased methionine incorporation acts to protect proteins from the reactive oxygen species. This process is just one highlight that demonstrates aaRSs connection to other metabolic roles.

Eukaryotic aaRSs & metabolism

A hallmark feature of eukaryotic aaRSs is the acquisition of appended protein domains. Relative to their bacterial homologs, the appended domains can be found throughout the molecule, attached to the N- or C-terminus and also insertions, although the former is more typical. Their functional roles have been unknown but are now being elucidated and include increasing aminoacylation efficiency by adding tRNA binding/specificity, creating interfaces for protein complex formation, and also serve other functions outside of translation. An example can be found with a C-terminal

appendage in human CysRS [15]. Comparing the activities of wild type CysRS and truncated variant lacking the domain, the presences of the appendage actually decreases the chemistry rate (k_{chem}), makes no change to k_{cat} , tightens $K_d[tRNA^{Cys}]$ and also adds nucleotide readout to the anticodon and the core [15]. Taken together, by slowing the chemistry rate and adding tRNA binding, the C-terminal domains improves the readout and specificity for the system resulting in increasing the quality of Cys-tRNA^{Cys} synthesis [15].

A ubiquitous feature in the eukaryotic translation apparatus is a large 2.5 MDa macromolecular complex containing a subset of aaRSs, noted as the multi-aminoacyl-tRNA synthetase complex (MARS) (Figure 6) [16-19]. This complex has been found

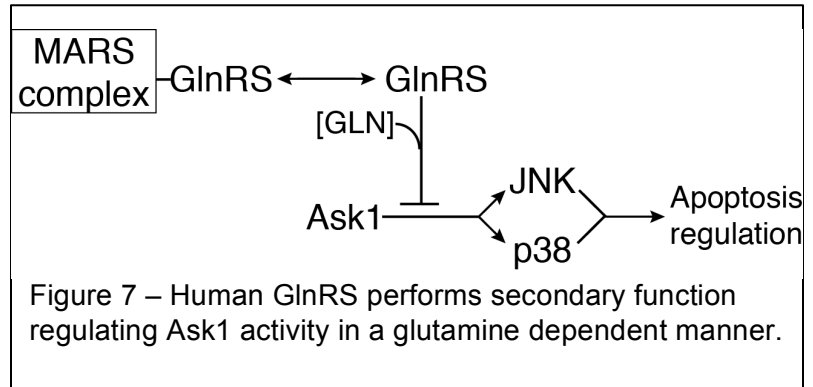
<i>S. cerevisiae</i> MARS		<i>T. brucei</i> MARS		<i>H. sapiens</i> MARS	
scaffold	aaRS	scaffold	aaRS	scaffold	aaRS
<u>Arc1p</u>	<u>GluRS</u>	Mcp1	AlaRS	Aimp1	AspRS
	<u>MetRS</u>	Mcp2	AspRS	Aimp2	ArgRS
		Mcp3	CysRS	Aimp3	GlnRS
			GlnRS		<u>GluProRS</u>
			<u>GluRS</u>		<u>IleRS</u>
			GlyRS		LeuRS
			IleRS		LysRS
			<u>MetRS</u>		<u>MetRS</u>
			ProRS		
			TrpRS		

Figure 6 – Three variations of eukaryotic multi-aminoacyl-tRNA synthetase complex. Underlined proteins are found in all three complexes.

in single-celled parasitic protozoa and humans. The hypothesis for the evolution of the MARS complex is based on the theme of eukaryotic cellular compartmentalization to help localize aaRSs near the ribosome and elongation factors. Furthermore it has been shown that for certain aaRSs, when in complex, increases aminoacylation efficiency relative to a free aaRS. For example in *Trypanosoma brucei*, $k_{cat}/K_m[tRNA]$ increases for MetRS 10-fold when associated in the MARS complex [20].

Another proposed role for the MARS complex has been hypothesized to act as a regulatory hub for aaRSs [19]. Most eukaryotic aaRSs have been implicated in functions

outside of translation. Again in humans, GlnRS is found in the MARS complex but also interacts with a mitogen-activated protein kinase (MAPK), apoptosis signaling



kinase 1 (Ask1) [21]. Ask1 regulates apoptosis by initiating the JNK and p38 pathways (Figure 7). Ask1 activation is sensitive to intracellular glutamine concentration [21]. Under high glutamine, Ask1 activity is turned off and this sensitivity is regulated by GlnRS. In this model, GlnRS acts as a glutamine indicator, where GlnRS can dissociate from the MARS complex, in turn binds glutamine where this binary complex can interact with Ask1, thus shutting down its downstream signaling pathway [21].

This GlnRS/Ask1 interaction has fundamental implications in cancer biology due to a unique role in glutamine metabolism. In a cancerous tissue, two of the most metabolized small-molecules are glucose and glutamine. In these aberrant cells, normal glycolysis and oxidative phosphorylation is not used for ATP synthesis, and the cell instead utilizes a unique aerobic glycolysis (also known as the Warburg Effect) [22-24]. Here cells produces ATP mostly through aerobic glycolysis, thus an enormous level of glucose is required to feed this process. This is obviously highly inefficient, an example of misregulated pathways in cancerous cells. Cancer cells use glutamine as a major carbon and primary nitrogen source to feed the citric acid cycle and nucleotide and amino acid biosynthesis pathways as these tissues continuously divide [22-24]. Thus in

healthy cells, a balanced interaction between GlnRS and ASK1 based on glutamine homeostasis maintains normal cell processes.

This is just one of many examples that demonstrates a crossroads between metabolic regulation and translation with implications in health. It ought not to surprise us given that aaRSs are thought to be ancient enzymes and essential genes. Here, Nature has selected them drawing upon the relatedness towards amino acid metabolism and their essentiality and adding another layer of functionality. This allows the cell to tune and balance protein synthesis, metabolism, and division, thus maintaining an overall homeostasis. Continued studies to understand how these enzymes are built will further our understanding in their role in aminoacylation as well as these non-canonical functions.

Synthetic biology

An emerging field with broad biological research implications is synthetic biology. Here scientists seek to develop tools through reengineering biological systems that will enable the creation of novel protein therapeutics, pathways for biofuel synthesis, biological microchips and diminish associated costs to synthesize antibiotics. Truly as this field continues to grow, only our scientific imagination limits the possibilities. One sub-discipline in synthetic biology is genetic code expansion by reengineering the translation apparatus to allow the incorporation of non-standard amino acids into protein synthesis. Incorporating non-canonical amino acids (ncAA) into protein synthesis has the potential to add new functionality to proteins. These include genetically encoded

fluorophores, residues to increase protein stability, adding site-specific bioconjugation target sites recapitulating post-translational modifications [25, 26].

In order to incorporate these amino acids into protein synthesis, synthetic biologists need to reengineer the natural genetic code with a new codon-amino acid pair [26]. This has been accomplished by changing the amino acid specificity of an aaRS-tRNA complex to recognize the ncAA and together with reassigning one of the three stop codons is then reassigned to represent the new ncAA.. Standard protein engineering – directed evolution methodology is used when mutating the amino acid binding pocket. During the screen, mutations are targeted to residues that directly interact with the amino acid substrate [26]. The selections occur by the read-through efficiency of a gene of interest with a stop codon within the open reading frame. Typically the target selection gene encodes an antibiotic resistance or a fluorescent protein.

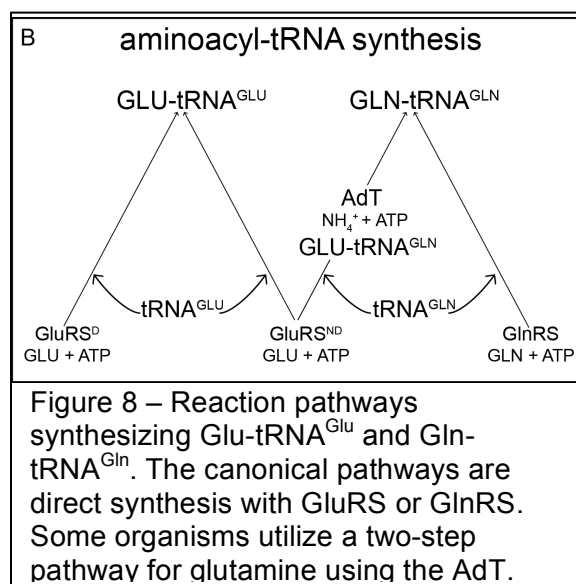
Two important aspects within these systems is the aaRS-tRNA scaffold choice, due to the plasticity of amino acid binding pocket, and tRNA orthogonality. Typical screens have been limited to mutating only a few residues within the substrate pocket, therefore choosing the best scaffold is important in order to achieve the desired specificity change [26]. In terms of the tRNA orthogonality, given that cell still require all of the endogenous aaRS-tRNA pairs for standard protein synthesis, the added engineered complex must work alongside it. Therefore, the added ncaaRS-tRNA pair must not interfere with the endogenous system. This has been resolved by taking advantage of poor cross species aminoacylation [26]. For example, adding an archaeal TyrRS-tRNA^{Tyr} pair into bacteria. For this scenario to be orthogonal the bacterial aaRSs cannot aminoacylate the added archaeal tRNA^{Tyr} nor can archaeal TyrRS interact with the bacterial tRNAs. At times this

requires further mutation to the tRNA to prevent cross charging, yet little has been done to monitor the kinetic effects from these tRNA mutations.

Two of the most successful reengineered systems for ncAA incorporation is TyrRS and PylRS [26]. Their relative success is likely due to their large amino acid binding pocket and ability to handle the tRNA anticodon mutation that accommodates stop codon complementarity. However these engineered systems are far from perfect and are limited due to poor read-through of the stop codon, poor tRNA specificity and inefficient amino acid incorporation [26]. These hurdles have been overcome, albeit haphazardly, by increasing the ncAA concentration in growth media and the use of multi-copy plasmids to increase tRNA and ncaaRS. Taking a step back to fundamentally understand these engineering endeavors, these studies are truly recoding the genetic code, thus we are asking a great deal to overcome billions of years of natural evolution at a systems level. Further work in understanding these systems and aaRS engineering methodology will be hugely beneficial, and will be discussed later.

Glutamine and glutamate phylogenetics

The studies in this work address all aforementioned concepts: coding, eukaryotic aaRS with their domains, and synthetic biology. The model system for the experiments stays within a single aaRS subfamily, the glutamyl- and glutaminyl-tRNA



synthetases (GluRS and GlnRS). This work expands upon a deep and rich understanding in macromolecular specificity, enzymology, molecular evolution and protein engineering concepts within the GluRS-GlnRS family.

GluRS and GlnRS both fall into the class Ib aaRS group, yet GlnRS is *not* universally conserved and is evolving in eukaryotes [27]. Organisms that do not code for a GlnRS utilize a two-step pathway using a tRNA-dependent amidotransferase (AdT) and a non-discriminating GluRSND for Gln-tRNA^{Gln} synthesis (Figure 8) [27, 28]. GluRSND is capable of synthesizing the canonical Glu-tRNA^{Glu} as well as Glu-tRNA^{Gln}. Obviously Glu-tRNA^{Gln} is detrimental to the cell and is a substrate for the AdT, which using a nitrogen source and ATP, synthesizes Gln-tRNA^{GLN}. Thus there are two GLU-GLN coding systems found throughout Nature (Figure 8).

We can reconstruct the evolutionary trajectory in the GluRS subfamily when considering the phylogenetic distribution of the GLU-GLN coding pathways (Figure 9) [27-29]. The last universal common ancestor utilized the two-step pathway using an AdT and GluRSND. From here two distinct GluRSND types, α and β class, where all bacteria and archaeal/eukaryotic are classified as such, respectively. These classes are differentiated by the structure of the anticodon binding domains; α -class use an α -helical cage and β -class have two stacked β -barrels. It is believed that LUCA was an α -class because of the relatedness between LysRS (I) and bacterial α -GluRSs.

As the progenitor eukaryotic/archaeal lineage split from bacterial, the β -GluRSND swapped the α -helical cage for the β -type anticodon-binding domain. All contemporary archaea still possess a β -GluRSND and AdT [27]. As stated, GlnRS arose in eukaryotes and is therefore a β -class GlxRS, evolving upon a duplication of a β -GluRSND. In eukaryotes, a fundamental coevolution mechanism takes place. As GlnRS evolves from

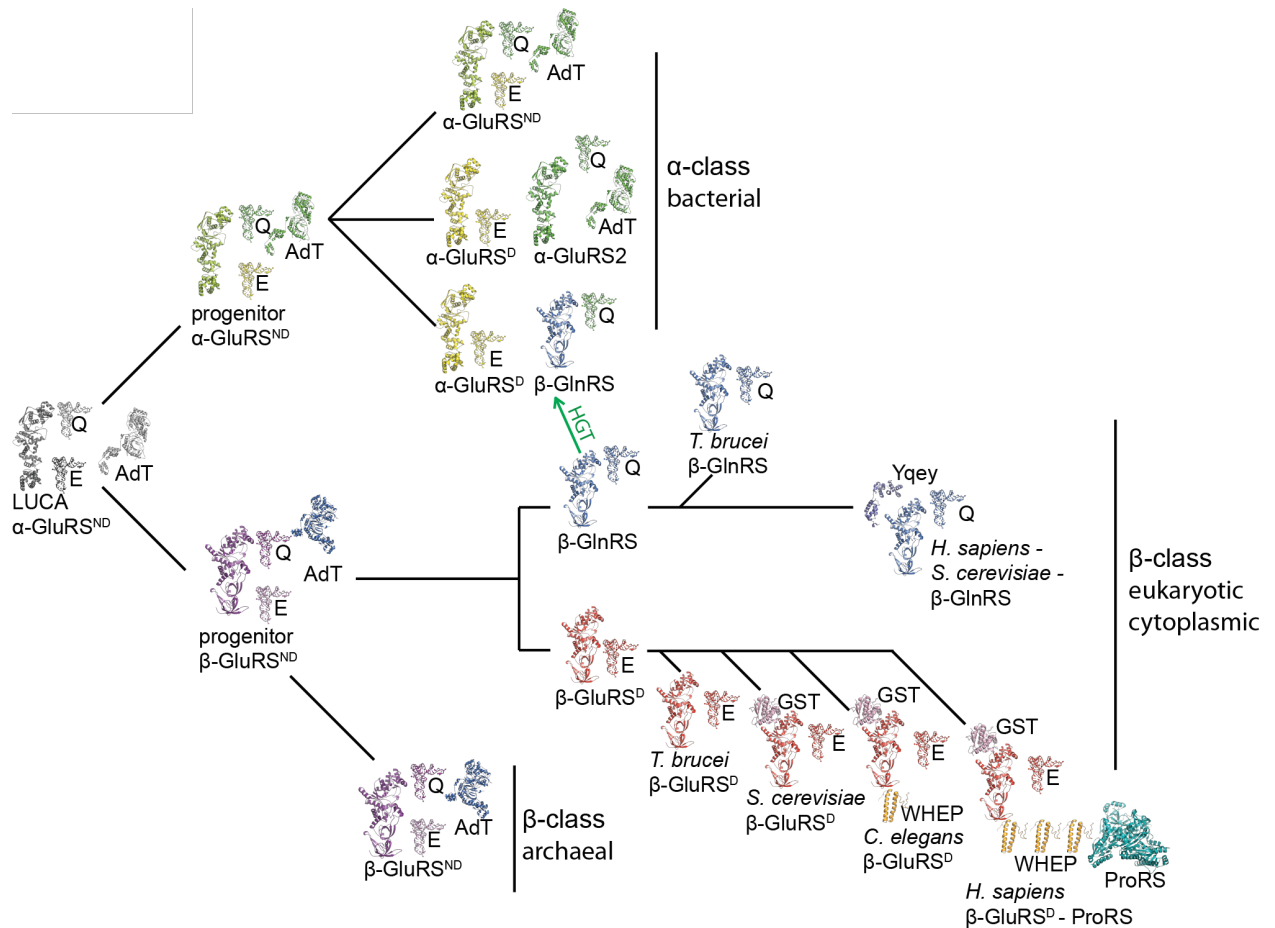


Figure 9 – Phylogenetic distribution of GlnRSs and GluRSs demonstrating which organisms possess a canonical GluRS and GlnRS or GluRSND and AdT. The bacterial horizontal gene transfer of the eukaryotic GlnRS is shown with a green arrow. Also in the eukaryotic lineage shows the various domain acquisitions and gene fusion events.

the GluRSND, it tightened the tRNA specificity while coevolving a new amino acid binding pocket for glutamine. Furthermore, it can lose the AdT gene, as it no longer requires this activity upon the acquisition of the more efficient direct Gln-tRNA^{GLN}

pathway via GlnRS. Furthermore, with the loss of the AdT and gain of GlnRS, the other GluRSND gene tightened its tRNA specificity for tRNA^{GLU} and maintained the glutamate specificity resulting in a discriminating β -GluRS^D.

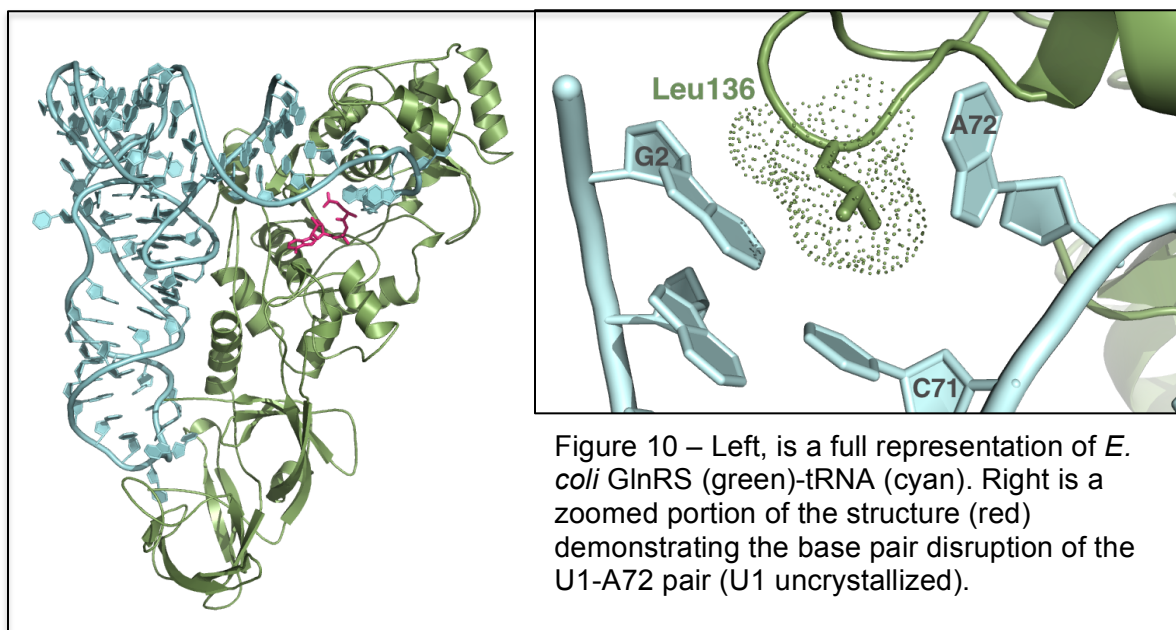
Most bacteria still use the α -GluRSND and AdT. However a subset of bacteria gained a eukaryotic β -GlnRS through a horizontal gene transfer event. Similar to what occurred in eukaryotes with the acquisition of GlnRS, the bacterial α -GluRS could tighten its tRNA specificity resulting in a discriminating α -GluRS^D-tRNA^{GLU} pair. However, some bacteria acquired a second α -GluRS gene, named GluRS2 [30]. In the small group of bacteria with two GluRS enzymes, one GluRS is specific for tRNA^{GLU}, while the GluRS2 is specific for tRNA^{GLN}. GluRS2 still retains glutamate amino acid specificity and is a *specific-misacylating* enzyme, thus necessitating the AdT. This leaves a total of three paths for Glu-tRNA and Gln-tRNA synthesis in bacteria (Figure 9).

As previously mentioned, eukaryotic aaRSs have acquired various appendages to their N- or C-terminus, GlnRS and GluRS are no exceptions [16]. GluRS gains a glutathione-S-transferase domain at the N-terminus and seen in lower eukaryotes [16]. As organism complexity increases, GluRS continues to gain domain complexity. These include the addition of small RNA binding domains, and in metazoan, GluRS is fused to ProRS (Figure 9). GlnRS gained a ~200 amino acid RNA binding appendage to the N-terminus in almost all eukaryotes [16, 31]. This domain is structurally homologous to the Yqey domain found in the AdT [32]. The functional roles for Yqey within the AdT are for tRNA^{Gln} specificity and discrimination *against* tRNA^{Glu} [33]. The fundamental aspect of the questions addressed throughout this work looks to understand what functional changes occurred to give rise to β -GluRS^D and β -GlnRS within eukaryotes. This

includes: the functional roles of the appended domains, the specificity determinants that generate a GlxRS-tRNA^{GLX} pair, and the mechanisms that dictate glutamate and glutamine specificity.

Mechanistic insight in GlnRS activities

Escherichia coli is a bacterial organism that acquired GlnRS. Coupling various crystal structures and mutational analysis using steady and pre-steady state kinetics provided the field with a deep understanding of substrate recognition in GlnRS [34-37]. The cocrystal of GlnRS-tRNA^{Gln}-Gln-AMP analog provided a structural framework for vital protein-substrate interactions. The structural details of the GlnRS and tRNA interactions led to the hypotheses for tRNA identity elements utilized for recognition. Clear interactions along the acceptor arm and anticodon were found leading to determine tRNA^{Gln} identity to be at positions: U1-A72, G2-C70, G3-C70, G10, C/U34, U35, and G36 [34, 38] (Figure 10). Two particular identity elements are worth noting when considering the phylogenetic trajectory of GlnRS. The U1-A72 base pair within the acceptor arm of tRNA^{Gln} is broken by an insertion loop, which likely could not be broken if the stronger pairing G-C were present (Figure 10) [34]. For anticodon position G36, an extruding loop within the anticodon binding domain makes clear hydrogen-bonding interactions along the Watson-Crick face [34]. These positions are of particular importance because they are distinct relative to tRNA^{Glu}, which has G1-C72 and C36. Since GlnRS arose from GluRS, that could recognize both tRNA^{Glu} and tRNA^{Gln}, evolving protein features to add tRNA^{Gln} specificity is an absolute requirement. tRNA^{Gln}



identity elements were further confirmed by the finding that mutating any of these ten RNA position greatly weakens tRNA specificity, based on a decreased $k_{cat}/K_M[\text{tRNA}]$ [38].

Comparing the apoenzyme against the tRNA bound structures of GlnRS, clear structural rearrangements are apparent upon binding [36]. However combining structural findings with detailed mechanistic kinetics can elucidate more indirect interactions between the two macromolecules. Taking advantage of transient state kinetics, one can probe how specific interactions affect elementary rate constants through catalysis [37, 39]. This includes directly measuring binding through K_D measurements as well the rate of chemistry. For example, as compared to wild type tRNA^{Gln}, the U35A mutation decreases k_{chem} 28-fold [39]. These are different from parameters obtained by steady state kinetics (k_{cat} and K_m), since steady state parameters characterize the total catalytic process. Mutating the tRNA anticodon that is 40 Å away from the active site and measuring microscopic parameters (as opposed to

steady state) demonstrates intramolecular communication. Combining steady and pre-steady state kinetics with structural work provides a powerful means to characterize protein-RNA recognition mechanisms.

As mentioned, GlnRS is one of the four class I aaRS that requires the tRNA for the first step in aminoacylation. The conformational changes that occur in GlnRS upon tRNA binding shows shifts within the amino acid binding pocket [34, 36, 40-42]. These observations begin to explain the tRNA dependence in amino acid activation since it orients the active site into a productive form. Furthermore, it is well documented of the existence of substrate binding interdependence between tRNA^{Gln} and glutamine [37]. Continuing with the anticodon mutation, U35A, mutating this identity element weakens K_d [GLN] and slows k_{chem} [37]. To further address the interdependence between tRNA^{Gln} and glutamine binding, a previous study systematically disrupted hydrogen bonding interactions along the entire protein-tRNA interface and measured glutamine binding [35]. The experiments found weakening the tRNA binding along the acceptor arm and anticodon position, a concomitant weakening in glutamine binding was observed, measured by K_d [GLN] using pre-steady state kinetics [35]. However, analogous mutations along the vertical arm actually tighten glutamine binding. Taken together, these findings draw a unique connection between tRNA and amino acid binding and show that substrate identities are linked.

With the clear importance the tRNA plays towards amino acid binding it would not be surprising to find direct interactions between the RNA and amino acid substrate, but this is not the case. Studying the glutamine binding pocket in GlnRS shows a clear lack of discriminatory interactions. The glutamine binding pocket is generated by relatively

ambiguous hydrogen bonding contacts and also uses bound water molecules to make substrate contacts [41-43]. Comparing structures of GlnRS bound with glutamine and near-cognate glutamate, both amino acids occupy the same space [44]. Since the only difference between glutamine and glutamate is a primary amine versus hydroxyl side chains, it was interesting to observe the clear lack of specificity. The observation was made describing the network of hydrogen bonding interactions were shifted where each H-bond acceptor became donors and vice versa; this phenomena is how glutamate could bind into glutamine pocket. As measured by steady and transient state kinetics, GlnRS is 10^7 times more specific for glutamine than glutamate [37]. To further highlight amino acid-tRNA interdependence, accompanied with the weakened amino acid affinity using glutamate, the tRNA specificity also weakens [37]. These findings leave the questions: what gives rise to this enormous amount of specificity? Again, what RNA and protein elements coevolved following the gene duplication event in eukaryotes that led to GlnRS and GluRS^D (Figure 9)?

Combining these previous studies creates a fundamental question as to how the genetic code translates glutamine-coding elements. It is clear GlnRS is highly specific for its substrates, and the selection for tRNA^{Gln} and glutamine is dependent upon each other. This interdependence is a fundamental mechanism ensuring cognate amino acid and tRNA are ligated together, safeguarding coding fidelity. After observing the ambiguous glutamine binding pocket and the poor capability to aminoacylate tRNA^{Gln} with glutamate, we sought to understand the basis of amino acid specificity in this class I aaRS.

Current works

Chapter 1 Synopsis: The glutamate and glutamine coding systems have been studied for over two-decades providing a rich understanding of the evolutionary history and mechanisms for molecular recognition. However, there is still much to be learned to further understand the coevolution of a β -GlxRS-tRNA^{GLX} pair, which this work seeks to address. Overall the bacterial systems are well characterized while the eukaryotic systems are poorly understood. This is unfortunate given that glutamine specificity arose within the eukaryotic lineage. These studies used *S. cerevisiae* GlnRS and GluRS as model enzymes to characterize the level of specificity each enzyme possesses for tRNA and amino acids. Both enzymes also utilize extra RNA binding domains. GlnRS has the N-terminal RNA domain, Yqey. GluRS interacts in *trans* with an RNA binding protein, Arc1p through a protein-protein interaction utilizing the N-terminal GST domain. Since both GlnRS and GluRS arose from a common progenitor β -GluRSND, this led to the hypothesis that perhaps the acquisition of the extra RNA binding domains assists in tRNA specificity, as each enzyme gained specificity for one tRNA^{GLX} (e.g. GlnRS gained tRNA^{Gln} specificity and lost tRNA^{Glu}.) However, in order to understand how eukaryotic GlxRS recognizes the tRNA substrate an understanding of what the tRNA identity elements are for glutamine and glutamate, which was not known. To be discussed, this work uncovers the eukaryotic tRNA^{Glu} and tRNA^{Gln} identity elements and elucidates how these identity elements interact with the cognate GlxRS. Furthermore a complete mechanistic characterization is carried out to ascertain the function for Yqey in GlnRS

and Arc1p with GluRS. Also addressed is the further characterization of the effects on tRNA identity elements in amino acid recognition. This work helps clarify the molecular signatures that coevolved between the β -GlxRS-tRNA^{GLX} pair resulting in GlnRS-tRNA^{Gln} and GluRS^D-tRNA^{Glu}. The project initially began through a collaborative project where I characterized the *S. cerevisiae* GlnRS using steady state kinetics and published in Grant et. al. 2012 *Nucleic Acid Research* [45].

Chapter II Synopsis: As previously mentioned, work in the *E. coli* GlnRS system uncovered the phenomena of substrate binding interdependence between glutamine and tRNA^{Gln}. This was related to the structural ambiguity of the glutamine binding pocket. To address this, we took a rational engineering approach to fundamentally change the specificity of our scaffold GlnRS-tRNA^{Gln}-Gln system towards glutamate specificity. Previous experimental work began finding the minimal protein mutations within the glutamine pocket to change the specificity to glutamate. It was found that amino acid specificity is distributed throughout the entire second half of the Rossmann fold, as it required a total of 22-mutations to completely change the specificity from glutamine to glutamate. These mutations expanded beyond the protein residues in direct substrate contact. As compared to naturally occurring GluRS activities, the mutated GlnRS (noted as a hybrid enzyme) was capable of efficiently aminoacylating tRNA^{Gln} with glutamate. These mutations recovered 16,000-fold specificity for glutamate. However, given amino acid and tRNA identities are linked, these studies found the tRNA specificity is substantially weakened, even after recovering amino acid specificity. In order to ascertain the success of the reengineering efforts, effective enzymatic benchmarks were needed. One of my first projects was to characterize an

archaeal GluRSND activities for these comparison purposes, which was published in Rodriguez-Hernandez et. al. 2010 in *Biochemistry* [46].

We hypothesized that in order to improve the specificity in our reengineered system, we must now recover the protein-RNA interface. Our approach was to take the scaffold tRNA^{Gln} and replace the glutamine RNA identity elements with tRNA^{Glu} identity elements. Applying our findings from the previous work characterizing the eukaryotic tRNA identity, we knew the tRNA^{Glu} identity elements. Furthermore, staying within the eukaryotic lineage is important, since it is where GlnRS arose. Briefly, these experiments uncovered tRNA identity elements, which represent a particular amino acid, also assists the enzyme in amino acid selection. This is a novel mechanism describing how the genetic code is translated – RNA identity functionally coupled with its coded amino acid through the aaRS.

This concept is of particular interest in synthetic biology and expanding the genetic code for ncAA incorporation. Researchers seeking to engineer orthogonal ncaaRS-tRNA-ncAA systems to incorporate ncAA into protein synthesis can utilize these results for rationally engineering methodology. Previously mentioned engineering ncaaRSs typically focus on mutations within the direct amino acid binding pocket. Our results highlight the distributive property of substrate recognition in aaRS, or at least in class I systems. These properties also expand to the tRNA substrate, as demonstrated by our hybrid GlnRS-tRNA system, to be discussed in detail in chapter II.

Addressed Questions

- What are the protein and RNA recognition elements in the eukaryotic GlxRS subfamily?
- Do RNA identity elements influence the specificity for its coded amino acid?

Chapter I

Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl-tRNA synthetases

This work has been published under this title in *The Journal of Molecular Biology* 2014
authored by Andrew Hadd and John Perona.

Introduction

Aminoacyl-tRNA synthetases mediate information transfer in all cells by specifically coupling amino acids with transfer RNAs (tRNAs) for protein synthesis.[1, 47] Minimal protein structures for the aaRS are generally found among the Bacteria and Archaea, while eukaryotic aaRS frequently possess additional domains that may assist in imparting novel functions.[48]; [49] For example, the WHEP domains of human glutamyl-prolyl tRNA synthetase (GluProRS) enable translational silencing,[50] while the UNE-S domain of SerRS confers nuclear localization and a role in transcriptional regulation essential for vascular development.[51] Other added domains are required for association of some of the enzymes to the large multi-tRNA synthetase (MARS) complex found in organisms ranging from trypanosomes to humans, which enhances protein translation by a channeling mechanism.[20, 52, 53] New aaRS domains may also modulate the kinetic properties of the enzymes for their canonical function in tRNA aminoacylation.[15, 32] This role is likely to be quite general, even if other functions are

also associated with the new domains, and is consistent with the general notion that tRNA and aaRS sequences have coevolved over evolutionary time.[3, 54] However, detailed appreciation of how the added domains of eukaryotic aaRS contribute to the catalytic mechanism of aminoacylation is still quite limited.[1]

A unique problem in aaRS coevolution is presented by the closely related glutamyl- and glutamyl-tRNA synthetase (GlnRS; GluRS) families. GlnRS was not present in the last universal common ancestor (LUCA);[27, 28, 55] instead, glutamine (Gln) was incorporated into proteins via a still-extant two-step pathway in which a nondiscriminating GluRS (GluRS^{ND}) first aminoacylates tRNA^{Gln} with glutamate (Glu).[56] In the second step, a tRNA-dependent amidotransferase converts the Glu- tRNA^{Gln} to Gln- tRNA^{Gln} via an ATP-dependent mechanism in which asparagine, Gln or ammonium ion serve as amide donor (Figure 8).[57, 58] All contemporary archaea, most eukaryotic organelles and some bacteria still retain this two-step pathway, which is mutually exclusive with the presence of GlnRS. GlnRS arose after the emergence of eukaryotes from the duplication and subsequent differentiation of the gene encoding GluRS^{ND} , and its presence in present-day bacteria is due to horizontal gene transfer (Figure 9).[8, 59]

In early eukaryotes, GluRS^{ND} also differentiated to generate the canonical discriminating GluRS^{D} enzymes. In this direction, the enzyme acquired additional discriminatory capacity to exclude efficient glutamylation of tRNA^{Gln} , while maintaining amino acid specificity for Glu. By contrast, in the direction of GlnRS, the enzyme instead differentiated to exclude tRNA^{Glu} as a substrate, while simultaneously undergoing a change in its amino acid substrate specificity from Glu to Gln. These

events are distinguishable from coevolution in most other aaRS systems, in which the enzyme and tRNA sequences evolve in the context of a fixed amino acid substrate identity, and the enzyme strictly maintains specificity for tRNAs in the same isoacceptor group.[3, 54] The requirement for coevolution of the amino acid binding site together with the enzyme-tRNA interface suggests that the determinants for amino acid specificity in the GluRS/GlnRS (GlxRS) family may reside partly in the RNA.[37, 46, 60, 61] In this context it is of interest that GluRS and GlnRS are among just four aaRS families (the others are ArgRS and class I LysRS) that require the presence of tRNA for synthesis of the aminoacyl adenylate reaction intermediate.[1]

tRNA identity determinants for bacterial GlnRS are well-established, and primarily comprise nucleotides at the acceptor end and anticodon loop.[38, 62-64] These elements have been rationalized in detail based on crystal structures of the *E. coli* GlnRS:tRNA^{Gln} complex.[65, 66] In addition to these parts of the molecule, bacterial GluRS appears to also recognize elements in the D-arm of its cognate tRNA^{Glu}, although these experiments are less definitive because they utilized weakly-performing tRNA transcripts lacking the crucial mnm⁵s²U34 wobble modification (there is no *E. coli* tRNA^{Glu} isoacceptor possessing the unmodified C34).[12, 67] Nonetheless, analysis of a large set of hybrid bacterial tRNAs intermediate in Glu and Gln acceptor activity established that the tertiary core region of tRNA^{Glu} functions as a negative determinant for GlnRS recognition, and that overlapping anticodon determinants between the two isoaccepting groups were avoided by focusing discrimination on the identity of the nucleotide at the anticodon 3'-position (G36 in tRNA^{Gln} and C36 in tRNA^{Glu}).[64]

Phylogenetic analysis has established that the ancestor of all GlxRS family members resembled contemporary bacterial GluRS in possessing an α -helical anticodon binding domain (in conjunction with the Rossmann fold catalytic domain characteristic of all class I aaRS) [68]. This ancestral GluRS was nondiscriminating and persists in many contemporary bacterial lineages that have not acquired GlnRS by horizontal transfer (Figure 9). In contrast, the archaeal GluRSND exchanged the α -helical cage for a pair of β -barrel domains, which are now found in all contemporary eukaryotic GluRS^D and GlnRS (Figure 9). Some insight into how the GluRSND to GluRS^D transition was accomplished in each structural context is available. In the bacterial α -helical class, a highly conserved arginine residue is responsible for recognition of C36 and is a determinant of GluRS^D specificity (in contrast GluRSND must accommodate both C36 in tRNA^{Glu} and G36 in tRNA^{Gln}). [69] In the archaeal β -barrel class, a conserved G36-recognizing surface loop in GlnRS is missing in GluRSND, and is also a determinant of specificity for the discriminating enzyme (in this case, the GlnRS). [61, 66, 68] Conserved differences in the structures of a protein surface loop recognizing the terminal 1-72 base pair of the acceptor stem are also important to distinguishing GluRSND from GlnRS. [61] However, the detailed trajectory of sequence evolution from GluRSND to GlnRS as it occurred in the early eukarya remains largely unknown. The complexity of the problem is highlighted by the finding that 23 amino acid mutations were required to convert the Gln binding site in *E. coli* GlnRS to permit efficient recognition of Glu, yet the resultant hybrid enzymes remained deficient in both k_{cat} and $K_M[tRNA]$ for the synthesis of Glu-tRNA^{Gln}. [46, 60] Other enzyme and RNA

determinants, as yet unidentified and probably located throughout the structure of the complex, are clearly also important.

Here we focus attention on a later stage of the coevolutionary process: the acquisition of additional domains and binding partners by eukaryotic GlnRS and GluRS^D. Eukaryotic GlnRS have acquired a roughly 200 amino acid N-terminal domain that is important for the proper assembly of the MARS complex.[70] The recently determined structure of this domain shows, remarkably, that it is homologous to the Yqey tRNA specificity-determining domain in the GatB subunit of the tRNA amidotransferase GatCAB – which specifically discriminates against tRNA^{Glu}. [32, 33] This suggests a role in tRNA discrimination that may contribute to coding specificity in eukaryotes. Yeast and bacterial GlnRS each poorly aminoacylate the cross-species tRNA^{Gln}, but fusion of Yqey to the *E. coli* enzyme improved its ability to aminoacylate yeast tRNA^{Gln}. [31, 71] This also suggests an important role for Yqey in coevolution. Interestingly, GlnRS of the bacterium *Deinococcus radiodurans* and related species has acquired the Yqey domain as a C-terminal appendage, where it functions to improve affinity for tRNA^{Gln}. [72]

Most eukaryotic GluRS have acquired an N-terminal GST domain. In yeast, this domain of the enzyme associates with the RNA-binding protein Arc1p in a ternary complex that includes MetRS. [73, 74] Arc1p improves the binding affinity of cytoplasmic GluRS to tRNA^{Glu}. [18, 75] Remarkably, yeast cytoplasmic GluRS can be targeted to the mitochondria where, in the absence of Arc1p, it misacylates mitochondrial tRNA^{Gln} with Glu as part of the two-step aminoacylation pathway in that

organelle.[76] Hence, Arc1p may function as a tRNA specificity factor and may also have an important role in sequence coevolution.

To elucidate the roles of the Yqey domain of eukaryotic GlnRS, and of the Arc1p binding partner of eukaryotic GluRS, we carried out detailed steady-state and transient kinetic studies of each enzyme from *S. cerevisiae*. We find that Yqey and Arc1p each strongly increase the affinities of the parent aaRS for their cognate tRNAs, but have little to no effect on the catalytic steps. Despite their acquisition at the same evolutionary stage at which coevolution away from GluRSND activity commenced, neither protein domain assists in enforcing specificity against noncognate tRNA. However, each new domain modulates the affinity of the parent enzyme for its cognate amino acid, a process that is coupled to the recognition of mutually exclusive tRNA recognition nucleotides primarily located in the acceptor stem and the 3' position of the anticodon. These functional aspects of contemporary eukaryotic GlnRS and GluRS appear to reflect their shared evolutionary history in diverging from a progenitor archaeal GluRSND.

Results

Eukaryotic GlxRS RNA binding domains modulate aminoacylation efficiencies

Contributions to aminoacylation efficiency made by the N-terminal Yqey domain of *S. cerevisiae* GlnRS, and by the Arc1p GluRS binding protein, were each examined by steady-state and transient kinetic analysis. For these experiments we utilized

unmodified tRNA transcripts synthesized *in vitro* by T7 RNA polymerase. In glutaminylation experiments conducted with *S. cerevisiae* GlnRS, we previously showed that k_{cat} and $K_{\text{M}}[\text{tRNA}]$ are identical for native *S. cerevisiae* tRNA^{Gln} and the unmodified transcript of the same primary sequence (Table 1).[32] To establish that an unmodified tRNA^{Glu(CUC)} transcript is a suitable GluRS substrate, we purified a mixture of native tRNA^{Glu(CUC)} and tRNA^{Glu(mcm5s2UUC)} from *S. cerevisiae* cells using affinity chromatography with a biotin-linked complementary oligodeoxynucleotide. *S. cerevisiae* tRNA^{Glu(CUC)} and tRNA^{Glu(mcm5s2UUC)} differ at just one acceptor-stem base-pair in addition to the difference at the wobble position, making full separation of the tRNA^{Glu(CUC)} isoacceptor impossible in the quantities required for this analysis. Kinetic analysis established that $k_{\text{cat}}/K_{\text{M}}[\text{tRNA}]$ is identical within experimental error for the native, post-transcriptionally modified tRNA^{Glu} mixture and the unmodified transcript (Table 1). We also note that the post-transcriptional modification of the similar mnm⁵s²U34 base at the wobble position of *E. coli* tRNA^{Glu} is crucial to aminoacylation efficiency *in vivo* and *in vitro*.[4, 6] This further dictated our choice to work with the tRNA^{Glu(CUC)} transcript. C34 is not known to be modified in any tRNA^{Gln} or tRNA^{Glu} species.

The *S. cerevisiae* GlnRS contains an N-terminal domain (NTD) that is conserved in eukaryotic enzymes and is not present in bacterial homologs. The NTD consists of 187 amino acids organized in two helical subdomains, and is followed by an unstructured 26-residue linker that links it with the main catalytic portion of the enzyme (CTD).[45] Computational modeling suggests that the NTD may bind the globular hinge region of the tRNA in a manner that is compatible with occupancy of the tRNA inner surface by

A. Kinetic parameters for steady-state and transient aminoacylation reactions							
	Steady state			Transient state			
	k_{cat} (s ⁻¹)	K_m [tRNA] (μM)	K_d [amino acid] ^a (mM)	$k_{obs,max}$ (s ⁻¹)	K_d [tRNA] (μM)	K_d [amino acid] ^a (mM)	Pre-steady state burst?
Arc1p-GluRS	1.6 ± 0.6	0.16 ± 0.03	39 ± 14	36 ± 7	0.23 ± 0.08	39 ± 1	yes
GluRS	1.5 ± 0.7	1.2 ± 0.3	20 ± 8	27 ± 7	4.2 ± 0.6	20 ± 3	yes
GlnRS	1.4 ± 0.2 ^b	0.19 ± 0.04 ^b	1.7 ± 0.2 ^b	22 ± 3	0.10 ± 0.03	5.9 ± 0.5	yes
ΔNTD-GlnRS	1.7 ± 0.2 ^b	5.6 ± 0.5 ^b	9.3 ± 0.3 ^b	23 ± 11	3.2 ± 0.6	34 ± 12	yes

B. Post-transcriptionally modified tRNA				
	Arc1p-GluRS		GlnRS	
	<i>in vitro</i> tRNA ^{Glu}	<i>in vivo</i> tRNA ^{Glu}	<i>in vitro</i> tRNA ^{Gln}	<i>in vivo</i> tRNA ^{Gln}
$k_{cat}/K_m[tRNA]$ (s ⁻¹ M ⁻¹)	1.0 ± 2.0 × 10 ⁷	1.9 ± 0.5 × 10 ^{7(c)}	0.7 ± 0.5 × 10 ^{7(b)}	1.2 ± 0.1 × 10 ^{7(b)}

Table 1 -

^aThe cognate amino acid was employed in all experiments.

^bThese data were previously reported (ref. 25).

^cThis parameter was determined under single-turnover conditions

the CTD (Figure 2).[77] Previously, we measured steady-state kinetic parameters for full-length *S. cerevisiae* GlnRS and an N-terminal deletion mutant lacking amino acids 1-186 corresponding to the NTD [(Gln4(187-809), here named ΔNTD]. These studies demonstrated that the tRNA^{Gln} and glutamine K_m values for ΔNTD were elevated by 30-fold and five fold, respectively, while k_{cat} for the two-step aminoacylation reaction was unaffected.[57]

To gain mechanistic insight, we carried out transient kinetics using molar excess of GlnRS over tRNA under conditions in which all three substrates (tRNA^{Gln}, ATP and Gln) were demonstrated to be saturating. These single-turnover reactions returned k_{obs} of 22 sec⁻¹ and 24 sec⁻¹ for wt GlnRS and ΔNTD, respectively, identical within experimental error, and roughly 10-fold higher than the value of k_{cat} (Table 1). In these experiments k_{obs} represents the rate of the combined chemical steps or of a closely linked first-order

conformational rearrangement.[78] To measure binding affinities for Gln and tRNA^{Gln}, we titrated the concentration of each of these substrates in single-turnover assays, while maintaining saturating concentrations of the nonvarying substrates. These measurements showed that the NTD contributes 28-fold and 7-fold, respectively, to tRNA^{Gln} and Gln binding affinities, values that are very similar to the effects observed on K_m values (Table 1). Taken together, these data are consistent with a rate-limiting step that follows the formation of product on the enzyme – most likely the release of Gln-tRNA^{Gln} into solution. The observation that K_m [Gln] is less than K_d [Gln] for both enzymes provides further corroboration of a late rate-limiting step. To further confirm this, we carried out reactions under pre-steady state conditions using concentrations of enzyme and tRNA that would allow detection of a rapid burst of product formation followed by a slower steady-state rate. Such a burst was indeed observed for both wt and Δ NTD GlnRS (Table 1 and Figure 11).

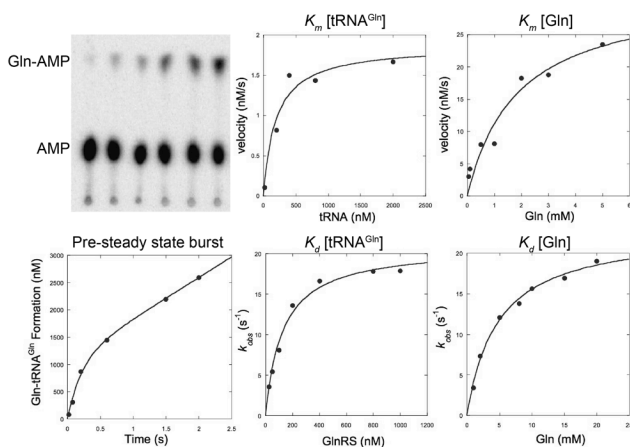


Figure 11 – Representative kinetics data for *S. cerevisiae* GlnRS. Upper left is a TLC showing product formation (Gln-AMP) compared to remaining substrate (AMP). This data was replotted as the pre-steady state burst (bottom left). This assay was also used to measure K_m , k_{cat} , K_d , and k_{chem} .

To compare the respective functions of the GlnRS Yqey domain and the Arc1p GluRS binding protein, we next performed steady state and transient kinetic analysis for *S. cerevisiae* GluRS using the tRNA^{Glu(CUC)} transcript as substrate. Under steady-state conditions, k_{cat} for glutamylation is unchanged by Arc1p, while K_m [tRNA] is reduced by

7.5 fold (Table 1). These effects are comparable to those observed for the influence of the Yqey domain on GlnRS, although the magnitude of the effect on $K_m[tRNA]$ is smaller. Interestingly, transient kinetics showed that the affinity (K_d) of GluRS for Glu is decreased two-fold by Arc1p, while the tRNA affinity is improved by over 20-fold (Table 1). Our finding with respect to the role of Arc1p in enhancing tRNA affinity is consistent with previous measurements by gel mobility shift.[75, 79] The affinity of GluRS for Glu is quite weak even without Arc1p, but is consistent with reports that intracellular Glu concentrations are as high as 90 mM in *S. cerevisiae*. [80] Arc1p also improves k_{obs} for glutamylation by about two-fold (Table 1). Effects of the protein on GluRS activity are thus complex: Arc1p promotes improved affinity for tRNA and first-order processing on the enzyme, while weakening Glu affinity. Burst analysis showed that the rate-limiting step is unchanged by Arc1p and (as in GlnRS) represents a late step after product formation on the enzyme.

Auxiliary protein domains do not contribute to tRNA specificity

Contemporary eukaryotic GlnRS and GluRS evolved from duplication of the gene encoding GluRSND in an early eukaryote (Figure 1).[27] The progenitor enzyme possessed the same nondiscriminating tRNA specificity as observed in the contemporary archaeal GluRSND. [61, 81] We therefore asked whether an inherent nondiscriminating activity might be retained within contemporary eukaryotic GlxRS. We hypothesized that the acquisition of the additional RNA binding domain (RBD) might provide each newly differentiated enzyme with a heightened level of specificity towards

the tRNA substrate. If this is indeed the case, removing the RBD could unmask residual activities characteristic of a GluRSND. This hypothesis was particularly enticing since *S. cerevisiae* cytoplasmic GluRS^D acts as a GluRSND when it is transported into the mitochondria.[76] The structural similarity of the GlnRS Yqey domain to the RBD in tRNA-dependent amidotransferases (AdT) is similarly suggestive, since the AdT discriminates against tRNA^{Glu} (in the context of Glu-tRNA^{Glu}).[32]

To test these hypotheses, we performed aminoacylation kinetics for both sets of enzymes using noncognate tRNA or amino acid as substrate, to examine formation of Glu-tRNA^{Gln}. Under transient state conditions with enzyme and substrate concentrations well above saturating levels for the cognate reactions (up to 400 mM Glu and 16 mM tRNA^{Gln} was used), no detectable synthesis of this product was observed for either GlnRS or GluRS, regardless of the presence of the RBD (Table 3). These are definitive findings given the very high sensitivity of the assay, which is capable of detecting extremely weak noncognate aminoacylations.[44] Indeed, the specificity of each *S. cerevisiae* enzyme is greater than that of *E. coli* GlnRS, which catalyzes Glu-tRNA^{Gln} synthesis with k_{cat}/K_m 10⁷-fold below its cognate reaction, and for which Glu-tRNA^{Gln} synthesis is readily detected at these concentrations of glutamate and tRNA.[44] Therefore, each cytoplasmic GlxRS-tRNA pair has fully lost the ancestral nondiscriminating activity in the course of coevolution, and the more stringent specificities arise from the conserved catalytic portions of each enzyme.

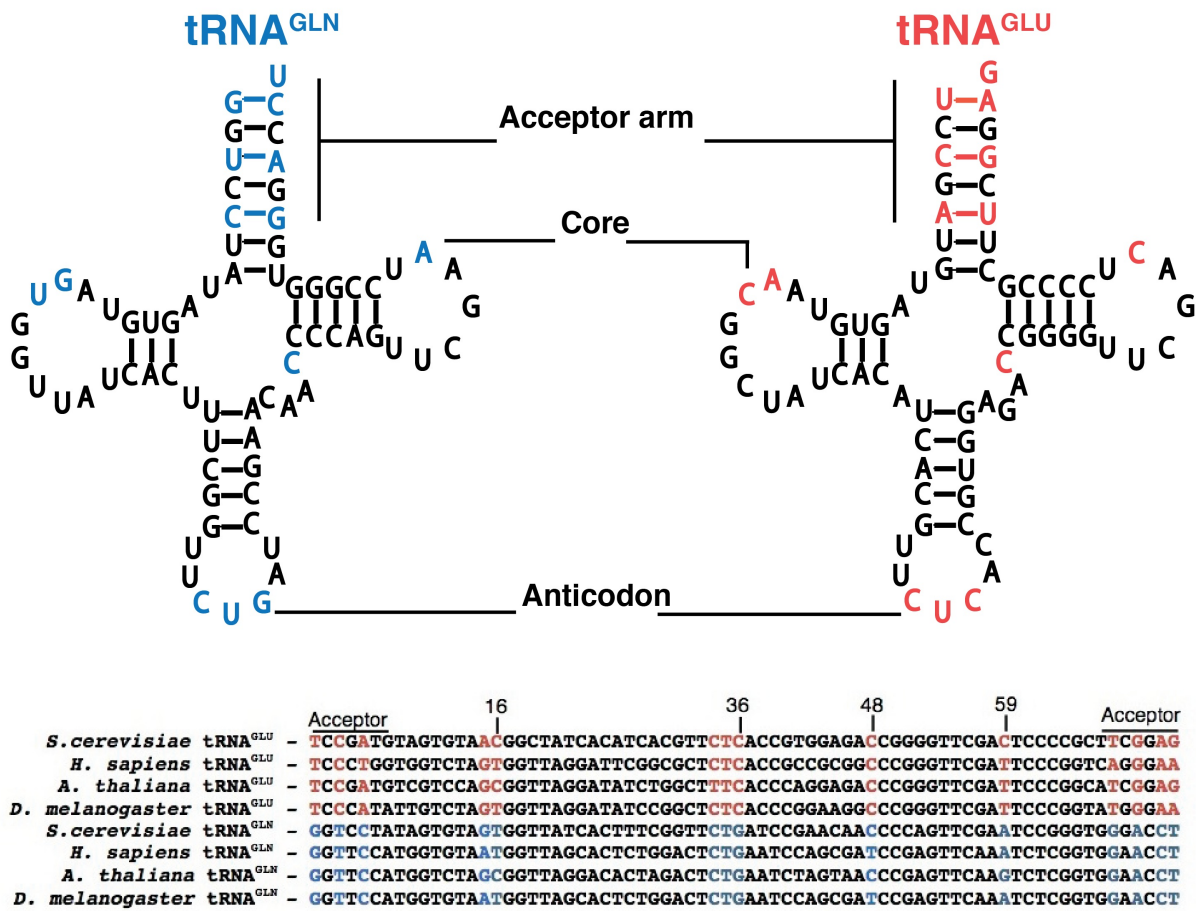
	GluRS-Arc1p		GlnRS	
	k_{cat}/K_m [tRNA] (s ⁻¹ μM ⁻¹)	k_{cat}/K_m [glutamate] (s ⁻¹ mM ⁻¹)	k_{cat}/K_m [tRNA] (s ⁻¹ μM ⁻¹)	k_{cat}/K_m [glutamine] (s ⁻¹ mM ⁻¹)
wt tRNA ^{Glu}	10	0.041	na	na
wt tRNA ^{Gln}	na	na	7.4	0.82
Qac36 tRNA ^{Glu}	0.028	0.0002	10.7	0.93
Eac36 tRNA ^{Glu}	0.232	0.0056	na	na

Table 2 - na – no activity was observed. Qac36 designates a tRNA^{Glu} in which acceptor arm (a) nucleotides, core (c) nucleotides, and nucleotide 36 from tRNA^{Gln} have been introduced (see Fig. 12). Eac36 designates a tRNA^{Gln} in which acceptor arm (a) nucleotides, core (c) nucleotides, and nucleotide 36 from tRNA^{Glu} have been introduced (see Fig. 12).

Recognition determinants for eukaryotic tRNA^{Glx}

The specific tRNA^{Glu} and tRNA^{Gln} nucleotides essential for cognate aminoacylation are unknown in eukaryotes. To gain insight into how these tRNAs are distinguished, we aligned a subset of sequences with a view towards discerning positions where nucleotide identities are mutually exclusive between the two isoacceptor groups (Figure 12). These alignments uncovered potential tRNA recognition elements for GlnRS and GluRS in the acceptor arm, at the 3' anticodon position and in the globular core (Figure 12). The consensus elements for Glu and Gln coding are distinct and are present at identical positions in each tRNA set. We thus hypothesized that they may function as both positive and negative aminoacylation determinants for the respective enzymes.

The proposed identity determinants are located at positions 73, 1-72, 3-70 and 5-68 in the acceptor arm, at the 3'-anticodon nucleotide (position 36), and at position 59 in the T-loop, which often pairs with nucleotide 16 near the top of the augmented D-stem (Figures 12). Eukaryotic tRNA^{Glu} species usually possess an unusual Py-Py pairing at this position (either C-C or U-U), while eukaryotic tRNA^{Gln} species almost exclusively feature the more common U16-A59 pair.[11] Nucleotides in this region often influence



enzyme binding via indirect readout: subtle modulation of the structure of the core region that is in turn recognized by the aaRS.[10] To maintain the integrity of the core domains in the yeast tRNAs that are the subject of these experiments, we also included the adjacent 15-48 base-pair in the portion of the core region that was mutated, although there is no distinction at this position for eukaryotic tRNA^{Glu} versus tRNA^{Gln} generally. The potential role of the core domain is of particular interest since the tRNA^{Glu} D-loop is an anti-determinant for recognition by the Yqey domain of

Methanothermobacter thermoautotrophicum AdT (GatDE).[33] We thus hypothesized that the Yqey domain in GlnRS might also function to discriminate against tRNA^{Glu}, while specifically recognizing these core identity elements in tRNA^{Gln}. This is consistent with the proposed location of the Yqey domain adjacent to the core region of tRNA, as deduced from structural modeling (Figure 13).[77] In the archaeal GatDE system, structural and functional analysis strongly suggests that the Yqey domain functions by excluding the unusually large tRNA^{Glu} D-loop based on steric constraints.[33] In contrast, the D-loops of yeast tRNA^{Glu} and tRNA^{Gln} are identical in size. Nonetheless, the possibility that Yqey could retain the same tRNA discriminatory function based on other structural distinctions between the yeast tRNAs remained.

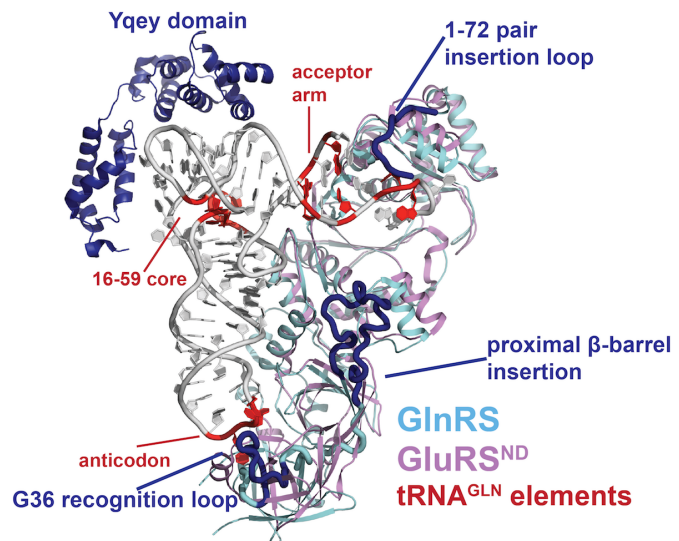


Figure 13 – Structural overlay of *S. cerevisiae* GlnRS (cyan) and *M. thermoautotrophicum* GluRSND (purple). The tRNA (white) is colored with important identity elements (red) (see Figure 12). In dark blue are GluRS specific insertions important for tRNA recognition.

We also considered the possibility that Arc1p might play a role in the recognition of specific tRNA nucleotides by GluRS. Biochemical reconstitution studies have previously shown that Arc1p exhibits broad tRNA binding specificity and interacts with a defined set of tRNAs.[82] When bound in a ternary complex with GluRS and MetRS, however, Arc1p strengthens the specific cognate tRNA binding affinities of each enzyme. Since the discriminating yeast GluRS^D differentiated from a GluRSND that also

bound tRNA^{Gln}, it seemed plausible that the acquisition of the Arc1p-binding GST domain of eukaryotic GluRS might have been driven in part by selective pressure to exclude tRNA^{Gln} binding.

To test these hypotheses, we generated a set of tRNA mutants as substrates for GluRS-Arc1p, GlnRS, and the respective enzymes lacking the eukaryote-specific RBD (Figures 12, 13). We first examined whether the consensus nucleotides act as positive determinants for the corresponding wt enzymes by transplanting each set into the noncognate tRNA framework. The 11 tRNA^{Gln} nucleotides U73, G1-C72, U3-A70, C5-G68, G36, G15, and U16-A59 were introduced into a tRNA^{Glu} transcript, generating the hybrid Qac36 tRNA^{Glu}, and the capacity of this tRNA for glutaminylation was analyzed. GlnRS is fully active toward this hybrid tRNA as measured by both $k_{cat}/K_m[tRNA]$ and $k_{cat}/K_m[Gln]$ (Table 2), and possesses no detectable activity toward wt tRNA^{Glu}. Interestingly, however, k_{cat} , $K_m[tRNA]$ and $K_m[Gln]$ are each increased by approximately 10-fold (Table 3). These data are consistent with our finding that the rate-limiting step for glutaminylation is product release (Figure 11). Considering $K_m[tRNA]$ as a proxy for the binding affinity of the aminoacylated tRNA product, the increase in the value of this parameter is consistent with weaker product binding and hence an enhanced dissociation rate that is reflected in the elevated k_{cat} . [39]

In the converse experiment, the 11 complementary tRNA^{Glu}-specific nucleotides were introduced into the tRNA^{Gln} framework (generating the hybrid Eac36 tRNA^{Gln}) for analysis of aminoacylation by GluRS-Arc1p. The GluRS activity was substantially reconstituted on this tRNA, although not fully: $k_{cat}/K_m[tRNA]$ and $k_{cat}/K_m[Glu]$ are 50-fold and seven-fold decreased, respectively, as compared to the analogous parameters

toward wt tRNA^{Glu} (Table 2). Most of the decreased efficiency arises from a lower k_{cat} , and no activity toward wt tRNA^{Gln} was detected (Tables 2, 3). The recovery of aminoacylation activity by adding determinants to the noncognate tRNA framework unambiguously shows the importance of those elements for the corresponding enzyme.

The accumulation of mutations in the ancestral tRNA^{Gln} sequence, at the 11 identified positions (Figure 12), would be expected both to improve aminoacylation by a coevolving GlnRS and to diminish aminoacylation capacity by a GluRSND coevolving on a trajectory towards GluRS^D. This concomitant use of positive and negative selections would allow the Glu and Gln tRNA identities to diverge from each other and would be

tRNA variant		k_{cat} (s ⁻¹)	K_m [tRNA] (μM)	k_{cat}/K_m [tRNA] (s ⁻¹ M ⁻¹)	K_m [amino acid] (mM)	k_{cat}/K_m [amino acid] (s ⁻¹ M ⁻¹)
GlnRS						
Q-ac36	tRNA ^{Gln}	1.4 ± 0.2	0.19 ± 0.04	7.3 × 10 ⁶	1.7 ± 0.2	820
	tRNA ^{Glu}	14 ± 3	1.3 ± 0.1	1.1 × 10 ⁷	15 ± 9	830
	tRNA ^{Glu}	5.0 ± 2	0.37 ± 0.08	1.3 × 10 ⁷	28 ± 18	180
Q-a_36	tRNA ^{Glu}	No activity				
ΔNTD-GlnRS						
Q-ac36	tRNA ^{Gln}	1.7 ± 0.2	5.9 ± 0.5	2.9 × 10 ⁵	9.3 ± 0.3	180
	tRNA ^{Glu}	6.7 ± 4	11 ± 4	6.1 × 10 ⁵	16 ± 9	420
	tRNA ^{Glu}	4.2 ± 2	10 ± 6	4.2 × 10 ⁵	95 ± 26	44
Q-a_36	tRNA ^{Glu}	No activity				
Arc1p-GluRS						
E-ac36	tRNA ^{Glu}	1.6 ± 0.6	0.16 ± 0.03	1.0 × 10 ⁷	39 ± 14	41
	tRNA ^{Gln}	0.17 ± 0.07	0.73 ± 0.2	2.3 × 10 ⁵	30 ± 19	6
	tRNA ^{Gln}	1.5 ± 0.6	2.9 ± 1	5.2 × 10 ⁵	60 ± 41	25
E-a_36	tRNA ^{Gln}	No activity				
GluRS						
E-ac36	tRNA ^{Glu}	1.5 ± 0.7	1.2 ± 0.3	1.3 × 10 ⁶	20 ± 8	75
	tRNA ^{Gln}	0.066 ± 0.02	1.8 ± 0.1	3.7 × 10 ⁴	9.8 ± 1.7	6.7
	tRNA ^{Gln}	0.34 ± 0.14	7.3 ± 1	4.7 × 10 ⁴	19 ± 2	21
E-a_36	tRNA ^{Gln}	No activity				

Table 3 - Qac36 designates a tRNA^{Glu} in which acceptor arm (a) nucleotides, core (c) nucleotides, and nucleotide 36 from tRNA^{Gln} have been introduced (see Fig. 12). Eac36 designates a tRNA^{Gln} in which acceptor arm (a) nucleotides, core (c) nucleotides, and nucleotide 36 from tRNA^{Glu} have been introduced (see Fig. 12).

reflected in contemporary sequences, as we have identified, in mutually exclusive nucleotide identities at the same positions in each tRNA. To test this notion we measured the capacities of GlnRS and GluRS to aminoacylate the hybrid Eac36-tRNA^{Gln} and Qac36-tRNA^{Glu} species, respectively. No activity of GlnRS toward the hybrid Eac36-tRNA^{Gln} was detectable, while GluRS activity toward Qac36-tRNA^{Glu} was decreased by 10^2 - 10^3 fold compared to wt tRNA^{Glu}, and neither tRNA^{Glu} nor Glu concentrations were saturable.

These data demonstrate that glutamate and glutamine identity elements indeed function as negative determinants against the opposing enzyme. However, the data also reveal that the systems are not perfectly symmetrical in this regard. The 11 recognition nucleotides do fully embody GlnRS discrimination against tRNA^{Glu}, since their exchange (i) fully reconstitutes GlnRS activity on the tRNA^{Glu} framework (Qac36 tRNA^{Glu}), and (ii) fully eliminates GlnRS activity on the tRNA^{Gln} framework (Eac36 tRNA^{Gln}) (Table 3). In contrast, our experiments with GluRS show that its activity on the exchanged Eac36 tRNA^{Gln} framework is not fully reconstituted, and its activity on the exchanged Qac36 tRNA^{Glu} framework is not fully eliminated. The retention of significant GluRS activity toward Qac36 tRNA^{Glu} suggests that some positive GluRS recognition determinants have not been identified, while the 50-fold decreased efficiency of GluRS toward Eac36 tRNA^{Gln} as compared to the cognate tRNA^{Glu} may further indicate the presence of unidentified negative determinants in tRNA^{Gln}. However, GluRS is decreased in aminoacylation capacity by over 10^7 -fold toward wt tRNA^{Gln} (Table 3). Therefore, the 11 nucleotide set that we have identified improves activity toward

tRNA^{Gln} by over 10⁵-fold, and must encompass a major portion of the full recognition set.

Roles of the tRNA core domains

To examine the roles of the Yqey domain and Arc1p protein in recognition of the tRNA^{Gln} and tRNA^{Glu} core domains, respectively, we constructed two additional hybrid tRNAs in which the three acceptor-stem base-pairs, the discriminator nucleotide at position 73, and anticodon nucleotide 36 were exchanged, but the core elements (nucleotide 15 and the 16-59 pair; Figure 12) were retained (generating two new hybrid tRNAs: Qa_36 tRNA^{Glu} and Ea_36 tRNA^{Gln}). For Gln coding, we then measured aminoacylation kinetics by wt GlnRS and Δ NTD GlnRS toward Q_a36 tRNA^{Glu} and the fully swapped Qac36 tRNA^{Glu}. The data reveal that wt GlnRS is generally insensitive to the tRNA core region with respect to $k_{cat}/K_m[tRNA]$, although there are compensating, roughly three-fold effects on k_{cat} and $K_m[tRNA]$ (Table 3). However, the three-fold decrease in k_{cat} is augmented by a two-fold elevation in the K_m for Gln when tRNA^{Glu} core region nucleotides are retained, so that $k_{cat}/K_m[Gln]$ is reduced by about five-fold. In the Δ NTD mutant, $K_m[GLN]$ is elevated to nearly 100 mM when Q_a36 tRNA^{Glu} is the substrate. These data suggest that the Yqey domain of eukaryotic GlnRS indeed recognizes the tRNA core region and communicates that recognition to the glutamine binding site some 20 Å distant.

We then performed the analogous experiments to examine whether Arc1p has a role in recognizing the core region of tRNA^{Glu}. We measured aminoacylation kinetics toward

Ea_36 tRNA^{Gln} and the fully swapped Eac36 tRNA^{Gln} using GluRS-Arc1p and GluRS alone. These data show that the core region nucleotides are not part of the tRNA recognition set for GluRS: indeed, glutamylation efficiency as measured by $k_{cat}/K_m[\text{Glu}]$ is three to four-fold greater when they are *not* exchanged into the tRNA^{Gln} framework, whether or not Arc1p is included in the reaction (Table 3). However, within the context of a lower overall efficiency (due to decreased k_{cat}), $K_m[\text{tRNA}]$ and $K_m[\text{Glu}]$ improve four-fold and two-fold, respectively, when tRNA^{Glu} core region nucleotides are included in the hybrid tRNA^{Gln}. Thus, core nucleotides may still act as minor identity elements. These effects are identical regardless of the presence of Arc1p. Therefore, Arc1p is not specifically involved in tRNA core region recognition, but exerts its effects more generally on tRNA binding. Consistent with this conclusion, we also find that $K_m[\text{tRNA}]$ for both hybrid tRNA substrates is improved by the same factor (about 2.5-fold) when Arc1p is present (Table 3).

Discussion

Our analysis substantially defines essential tRNA recognition nucleotides for eukaryotic GlnRS and GluRS. Key nucleotides in both systems, demonstrated by transplantation into the opposing noncognate tRNA framework, are located at acceptor stem nucleotides 73, 1-72, 3-70 and 5-68, and at the 3'-anticodon position 36. The identities of the nucleotides at these eight positions are mutually exclusive in most eukaryotic tRNA^{Gln} and tRNA^{Glu} species. We suggest that this finding for the contemporary aaRS-tRNA pairs is a vestige of the evolutionary process by which an

ancient β -class GluRSND in an early eukaryote was able to aminoacylate both species (Figure 9). After the gene duplication, each enzyme/tRNA pair differentiated such that noncognate aminoacylations were excluded. Key changes in the protein structures suggested to be important in this process are the acquisition of surface loops in GlnRS that interact at the 3'-anticodon position and adjacent to position 73 and the 1-72 base-pair, each of which is missing in an archaeal GluRSND (Figure 13).[61, 68] Our finding that these positions in eukaryotic tRNAs are central to their identities is thus consistent with the evolution of the GlnRS and GluRS proteins. Such evolution, beginning from the same ancestral framework, presumably occurred together with the differentiation of the tRNAs at complementary, nearby positions of the macromolecular interface. A crystal structure of a eukaryotic GluRS^D, presently unavailable except for the Arc1p interaction domain,[83] would be helpful in further studying the evolutionary trajectory. We also note that the pathways for evolutionary development likely include the acquisition of epistatic mutations that generate restrictive structural constraints, thereby limiting the capacities of the eukaryotic enzymes to reacquire their ancestral activities.[84]

This set of eight recognition nucleotides in eukaryotic tRNA^{Gln} fully embodies all the necessary discrimination for GlnRS against tRNA^{Glu}, since k_{cat}/K_m toward Qac36tRNA^{Glu} is reconstituted to the same level as cognate tRNA^{Gln}, and no residual activity can be measured toward Eac36tRNA^{Gln} (k_{cat}/K_m decreased by at least 10^7 -fold) (Table 3). In bacteria, a full GlnRS recognition set sufficient for full discrimination against tRNA^{Asp} encompassed 15 positions: G73, U1-A72, G2-C71, G3-C70, G10, and all seven anticodon nucleotides.[63] Even this much larger set was insufficient, however, for transplantation of GlnRS recognition nucleotides into the bacterial tRNA^{Glu}

framework.[64] Since both extremities of the tRNA are fully replaced in the 15-nucleotide set, it is likely that the bacterial tRNA^{Glu} core region functions as a negative determinant against bacterial GlnRS. This represents a significant distinction in the specificity determinants for glutamine coding in bacteria *versus* eukaryotes.

While introduction of the eight yeast tRNA^{Glu} recognition nucleotides 73, 1-72, 3-70, 5-68, and 36 into tRNA^{Gln} reconstituted yeast GluRS activity by over 10⁵-fold, activity remained diminished by 50-fold compared to cognate tRNA^{Glu}. Therefore, differentiation of the ancestral tRNA^{Gln} and tRNA^{Glu} sequences for discrimination by distinct enzymes also included further elements beyond the mutually exclusive eight positions - most likely the acquisition of additional negative determinants in tRNA^{Gln}, and the coevolution of further positive recognition nucleotides in tRNA^{Glu}, for GluRS. In bacteria, 12 nucleotides in bacterial tRNA^{Glu}, when introduced into the tRNA^{Gln} framework, substantially reconstituted activity (albeit within an experimental context in which K_m [tRNA] values were elevated by about 50-fold due to the lack of the U34 modified base).[12] A number of substitutions within the core region of bacterial tRNA^{Glu}, at positions U11-A24 and U13-G22-A46, were required for this. By contrast, as we observed for yeast tRNA^{Gln}, the core region of yeast tRNA^{Glu} does not play a role in recognition by GluRS or discrimination by GlnRS (Table 3). Again, this provides a strong contrast between the bacterial and eukaryotic coding systems for glutamate.

These distinctions likely reflect the much greater divergence of the bacterial coding systems, in which GluRS is of the opposite a structural class, and the evolution of GluRSND to GluRS^D upon invasion of GlnRS occurred in the context of marked differences in many of the parallel aminoacylation systems. Sequence relationships

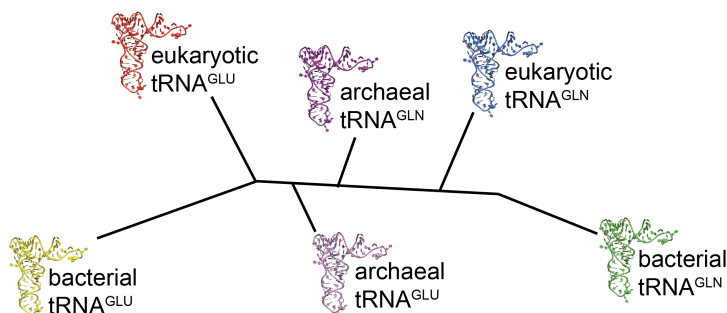


Figure 14 - Depiction of the evolution of tRNA^{Glu} and tRNA^{Gln}. The figure is based on a maximum likelihood analysis as previously reported, 54.

among tRNA^{Gln} and tRNA^{Glu} species across the three domains of life were previously elucidated using all available sequences in the tRNA database,[85] and are also informative in this context (Figure 14).[61] The analysis

shows that archaeal tRNA^{Gln}/tRNA^{Glu} pairs are most similar to each other, probably reflecting their aminoacylation by the same GluRSND in each organism. Eukaryotic tRNA^{Glu} and tRNA^{Gln} sequences are more greatly divergent, but retain some similarities perhaps because they coevolved together with enzymes that arose from the same GluRSND. Bacterial tRNA^{Glu} and tRNA^{Gln} are the most divergent pair. This is presumably due to the fact that their sequence coevolution occurred after a GlnRS horizontal gene transfer that required accommodating a highly distinctive new protein into the coding apparatus.

Our experiments also demonstrated that the yeast Yqey and Arc1p protein domains presently do not function to differentiate the Gln and Glu coding systems. The primary functions of Yqey and Arc1p are to increase the affinities of their respective enzymes for cognate tRNA and amino acid (Table 1), but neither protein increases specificity against noncognate tRNA or is significantly involved in tRNA sequence readout (Tables 2, 3). One possible explanation of this finding could be that the Arc1p-binding N-terminal GST module of eukaryotic GluRS, and the N-terminal Yqey domain of GlnRS, were relatively late additions that did not originate based on selective pressure to improve coding

specificity. Instead, the Yqey domain of GlnRS may have been acquired to facilitate the correct assembly of the MARS complex; in humans, its deletion generates a modified particle that, while retaining GlnRS, no longer accommodates arginyl-tRNA synthetase (ArgRS) or the auxiliary protein p43.[70] In contrast, the Arc1p-binding GST domain of yeast GluRS may have arisen as part of a metabolic control mechanism. Under fermentation conditions, GluRS is associated with Arc1p in the cytoplasm. However, when a fermentable carbon source is limiting, the cell shifts towards respiration. Arc1p then releases GluRS and a fraction of the enzyme is transported into the mitochondria to support the increased rate of mitochondrial protein synthesis.[76] Interestingly, the two-fold improved affinity of GluRS for Glu in the absence of Arc1p (Table 1) is consistent with a drop in intracellular glutamate concentrations from 90 mM to 40 mM during this shift, perhaps allowing the portion of the enzyme that remains in the cytoplasm to maintain efficient glutamylation.[80] The transported GluRS misacylates mitochondrial tRNA^{Gln} with Glu, and a second distinct GluRS^D is imported to catalyze Glu-tRNA^{Glu} synthesis in the organelle.[76] Cytoplasmic tRNA^{Glu} and mitochondrial tRNA^{Gln} share a common U1-A72 pair, while mitochondrial tRNA^{Glu} possesses an unusual G1-U72 wobble pair. This may provide part of the rationale for why the yeast mitochondria retains two GluRS enzymes rather than a single GluRSND.

Although the Yqey domain of contemporary yeast GlnRS does not provide specificity against tRNA^{Glu}, our data do not rule out the possibility that Yqey did play this role at an early evolutionary stage in eukaryotes, and that it was later retained because it became useful in eukaryote-specific functions such as assembly of the MARS. Both the ancestral GlnRS and the archaeal GatDE amidotransferase clearly had an acute

requirement for a high level of specificity against the tRNA^{Glu} scaffold. In archaea, this specificity is accomplished in significant part by means of an enlarged tRNA^{Glu} D-loop that excludes GatDE binding through its Yqey domain interface.[33] We speculate that early eukaryotes differentiating away from the archaeal lineage may also have featured archaeal-like tRNA^{Glu} species with enlarged D-loops, so that the ancestral eukaryotic GlnRS would have clearly benefited by recruiting Yqey to play the same role as in GatDE. Later coevolution of GlnRS and tRNA^{Gln} sequences at the extensive bacterial-like protein-RNA interface, together with shrinkage of the enlarged tRNA^{Glu} D-loop as this species coevolved with GluRS, may have rendered the Yqey domain unnecessary in tRNA sequence discrimination as we have observed. The retention of Yqey in contemporary GlnRS would then reflect its new role in facilitating formation of the MARS, or possibly in other cellular functions unrelated to translation.[70]

It is also of interest that bacterial GluRS do not possess Yqey as an N-terminal extension, suggesting that lateral transfer of Yqey-encoding eukaryotic GlnRS genes was deleterious to the recipient cell. This may have occurred because the presence of Yqey renders glutaminylation of the distinctive bacterial-type tRNA^{Gln} inefficient. Early experiments indeed showed that yeast GlnRS cannot efficiently glutaminylate bacterial tRNA^{Gln}, although no kinetic parameters were determined.[31] A more thorough study of this question using bacterial and yeast deletion and fusion proteins, with bacterial tRNA^{Gln} as substrate, could shed some light on this question. We speculate that the distinctive nature of bacterial tRNA^{Gln} could have served as a barrier for entry of a Yqey domain as an N-terminal extension to bacterial GlnRS – just as the distinct archaeal tRNA^{Gln} appears to block acquisition of any GlnRS enzyme in archaea.[61]

Finally, we also note that the interdependence of tRNA and glutamine binding that is well-documented in *E. coli* GlnRS is also present in the yeast GlnRS enzyme.[37] In this case there is a more prominent role for the tRNA core domain: core region nucleotides at positions 15-48 and 16-59 in the augmented D-stem are specifically read out by the Yqey domain, presumably in conjunction with other interactions made by the catalytic domain, and generate an approximately four-fold effect on the complementarity of the distant active site for Gln (Table 3). Thus, allosteric pathways bridging tRNA and amino acid binding sites are present within the eukaryotic GlnRS complex, and these are modulated by the Yqey domain. It is also worth noting that, while the introduction of just eight nucleotides outside the core region fully embodies tRNA^{Gln} identity at the level of $k_{\text{cat}}/K_m[\text{tRNA}]$ (Table 3), the individual k_{cat} and K_m parameters are elevated by ten-fold and seven-fold, respectively, and the $K_m[\text{Gln}]$ is also substantially weakened. Although the effects are smaller, core region nucleotides of tRNA^{Glu} are similarly capable of modulating $K_m[\text{Glu}]$, in a manner that is independent of the presence of Arc1p (Table 3). Thus, the core regions of both tRNAs assist the corresponding enzymes in amino acid coupling without playing a role in tRNA discrimination.

Chapter II

Recoding aminoacyl-tRNA synthetases for synthetic biology by rational protein-RNA engineering

This work has been published under this title in *ACS Chemical Biology* 2014

authored by Andrew Hadd and John Perona.

Introduction

The expansion of genetic codes through incorporation of noncanonical amino acids (ncAAs) offers substantial potential to develop designer proteins capable of yielding new insights into cellular processes.[25, 26] Engineering of the translational apparatus is also a key foundational technology in synthetic biology, with possible applications to metabolic pathway reconstruction, design of antimicrobial resistance, cancer therapeutics, and other areas.[45, 86] These aspirations depend on the creation of novel aminoacyl-tRNA synthetases (aaRS) capable of ligating one or more ncAAs to a new tRNA that reads an unused codon triplet or quadruplet – most commonly the UAG amber stop codon.[87-90] A directed evolution approach involving both positive and negative selections, applied to libraries targeting an aaRS active site, has allowed incorporation of over 100 ncAAs in bacterial or eukaryotic cells.[88, 91, 92] The choice of which aaRS scaffold to employ in these selections has been largely dictated by the

requirement for orthogonality: neither the new aaRS nor the new tRNA may substantially cross-react with endogenous parallel aaRS-tRNA systems. In some cases orthogonality may also require that similar positive and negative selections be carried out to optimize the tRNA sequence.[93, 94]

Although this directed evolution approach has been successful in producing ncAA-containing proteins *in vivo*, better integration of the new aaRS-tRNA pairs into the translational apparatus requires further efforts. Very high concentrations of ncAAs and overexpression of tRNAs are presently needed for efficient protein synthesis, but these features diminish cellular fitness by increasing competitions with endogenous aaRS for both amino acid and tRNA pools.[89, 95] Further, limits to the number of amino acid positions that can be sampled in libraries have generally confined explorations of aaRS structures to the immediate amino acid binding site environments. Limitation of the selections to a small number of aaRS scaffolds, primarily methanogen tyrosyl- and pyrrolysyl-tRNA synthetases (TyrRS; PylRS), likely further limits the range of ncAAs that can ultimately be incorporated.[90, 96, 97]

Weak activities of genetically selected aaRS in amino acid activation and tRNA transfer,[89, 90, 98] and susceptibilities of enzyme-bound noncanonical aminoacyl adenylates to competitive attack by water,[99, 100] underlie the requirements for high concentrations of ncAAs and orthogonal tRNAs in selections. We suggest that these deficiencies in the catalytic performance of aaRS enzymes emerging from directed evolution may be remediated through methodologies that incorporate rational design.[89] This approach is facilitated by reliable crystal structures bound to cognate amino acid for nearly all of the 24 phylogenetically independent families of aaRS.[1]

The aaRS families are partitioned into two structural classes (class I and class II), possessing class-specific homologous catalytic domains. Within each class, the exhaustive sequence information for aaRS (and tRNAs) then allows construction of structure-based alignments and predictions of selectivity determinants for coded amino acids. An early application of this rational design approach to a different enzyme family was the successful identification of chymotrypsin specificity determinants by transplantation of amino acids into the homologous trypsin scaffold.[101] We posited that rational design could be similarly effective in elucidating how homologous aaRS architectures have differentiated to select among the large number of amino acids present in the cell. This approach is analogous to the transplantation of proposed tRNA identity nucleotides into noncognate tRNA frameworks, which is long established in the aaRS field.[3] Insights derived from this rational engineering should then assist in formulating new approaches to obtain novel aaRS capable of encoding ncAAs for synthetic biology.

We have chosen *Escherichia coli* glutamyl-tRNA synthetase (GlnRS) as a model class I enzyme for rational design based on extensive structural and mechanistic studies, and on its function as a relatively small 553 amino acid monomer (Figure 1A).[35, 65] GlnRS requires tRNA^{Gln} as a cofactor for synthesis of the activated glutamyl adenylate intermediate, and is thus properly considered as a ribonucleoprotein (RNP) enzyme.[1] Previously, we reported early steps in a multistage design process in which amino acids derived from human glutamyl-tRNA synthetase (GluRS) were introduced into the GlnRS scaffold.[46, 60] GlnRS originated in early eukaryotes from the duplication and subsequent differentiation of a gene encoding a

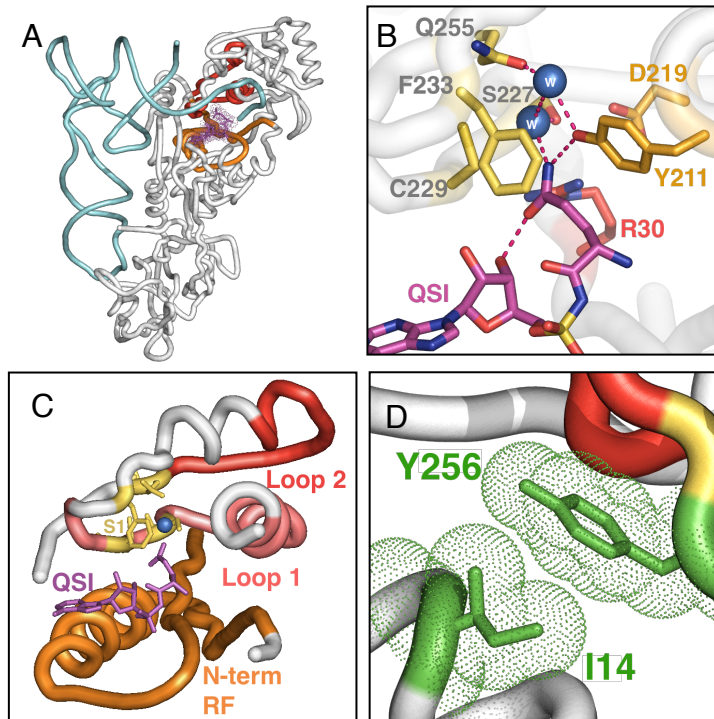


Figure 15 - A) Ribbon representation of the *E. coli* GlnRS-tRNA^{Gln} complex. The tRNA is in light blue. The N-terminal and C-terminal portions of the catalytic Rossmann fold (RF) are depicted in orange and red, respectively. An analog of glutaminyl adenylate is shown bound in the active site. B) Depiction of the hydrogen bond interactions (dotted lines) made with the glutamine substrate within the immediate binding pocket (S1). Residues Q255, F233, S227 and C229 were replaced to generate the S1 mutant (R30, Y211 and D219 are conserved in eukaryotic GluRS). Two interstitial water molecules are shown as blue spheres. C) Depiction of the RF bound to the glutaminyl-AMP analog QSI. The structurally conserved N-terminal portion of the RF (orange) forms most of the ATP binding interactions. The two exchanged loops in the L1L2 variants are shown in pink and red, respectively. The full list of mutations is provided in Table 1. D) van der Waals contacts between Tyr256 in L2 and the distal Ile14 residue, preserved in the GlnRS L1L2 W256Y hybrid enzyme.

nondiscriminating archaeal GluRS (GluRSND) capable of aminoacylating both tRNA^{Glu} and tRNA^{Gln} with glutamate (Glu).[59] The presence of GlnRS in some contemporary bacterial taxa is solely a consequence of horizontal gene transfer from eukaryotes, and the enzyme is more closely related to eukaryotic than bacterial GluRS.

All GluRS possess two arginine (Arg) residues that bind the Glu substrate carboxylate group, while all GlnRS retain only one Arg (Figure 15B; Figure 6). We previously showed that introduction of the second Arg into GlnRS (GlnRS C229R) lowers $K_m(\text{Glu})$ from >750 mM to 240 mM, but without improving k_{cat}/K_m for Glu-tRNA^{Gln} synthesis (which is diminished by 10^7 -fold

compared to Gln-tRNA^{Gln} synthesis by wt GlnRS (Table 1)).[60] Most GlnRS and GluRS also possess three additional distinguishing residues in their homologous primary amino acid binding sites (S1 sites). Further introduction of these three GluRS residues into GlnRS generated the S1 GlnRS enzyme (Table 4; Figure 15B), which exhibited 20-fold improved k_{cat}/K_m for Glu-tRNA^{Gln} synthesis compared to WT GlnRS. However, S1 GlnRS remains highly inefficient, with k_{cat}/K_m still reduced by nearly 10⁶-fold compared to Gln-tRNA^{Gln} synthesis by wt GlnRS.[60] This full replacement of all first-shell protein residues represents the first stage of the rational design effort. It demonstrated that determinants of amino acid selectivity in GlnRS are primarily nonlocal. Neither S1 GlnRS nor the enzymes produced in the later design stages retain any detectable activity for Gln-tRNA^{Gln} synthesis.

In the second stage of design, we examined more distal elements of protein structure within the second half of the catalytic Rossmann fold (RF), which is primarily responsible for amino acid binding in class I aaRS (Figure 15C). Because protein sequence alignments did not identify significant conserved differences between GlnRS and GluRS enzymes outside of the S1 site (Figure 16), we swapped distal human

	30	208	Loop 1	Loop 2	256
<i>E. coli</i> GlnRS	R/	/YPMYDF	THCISDALEGITHSLCTLE	FQDNRLYD	WVLDNITIP-VHPRQY
<i>H. sapiens</i> GlnRS	R/	/YPTYDY	THCLCDSIEHITHSLCTKE	FQARRSSYFW	LCNALDV--YCPVQW
<i>S. cerevisiae</i> GlnRS	R/	/YPTYDF	THCLVDSMENITHSLCTTE	FYLSRESYEW	LCDQVHV--FRPAQR
<i>T. brucei</i> GlnRS	R/	/YPSYDY	THCLIDSLENIDYSLCTLE	FETRRESYFW	LLEQLNL--WRPHVW
<i>H. sapiens</i> GluProRS	R/	/YPTYDF	ACPIVDSIEGVTHALRTTE	YHDRDEQFYW	IEALGI--RKPYIW
<i>S. cerevisiae</i> GluRS	R/	/YPTYDF	CVPIVDAIEGVTHALRTIE	YRDRNAQYDW	MLQALRL--RKVHIW
<i>T. brucei</i> GluRS	R/	/YPTYDF	CCPIIDSVEGITHALRTNE	YHDRNDQYYW	FCDALGL--RKPIVE

Figure 16 - Alignment of selected GlnRS and eukaryotic GluRS sequences in the amino acid substrate binding site regions. Residues in the first shell, immediately adjacent to the bound amino acid, are indicated in yellow. The tRNA-binding dipeptide at positions R237-R238 in *E. coli* GlnRS is colored blue.

GluRS peptides into S1 GlnRS based on structural considerations. Replacement of two surface loops bridging RF secondary structure elements (loops L1 and L2; Figure 15C) improved $k_{\text{cat}}/K_{\text{m}}$ for Glu-tRNA^{Gln} synthesis by 10³-fold compared to S1 GlnRS, and reduced $K_{\text{m}}[\text{Glu}]$ by 40-fold (L1L2 GlnRS; Table 4).[60] Replacement of distal surface loops in serine proteases similarly effected conversion of amino acid selectivity in the context of peptide bond hydrolysis, providing inspiration for these experiments.[100]

In the third design stage, we examined all interactions made by the 23 replaced amino acids in L1L2 GlnRS with surrounding regions of the protein structure.[46] This allowed assessment of the role of selected third-shell residues in amino acid substrate discrimination, and of the potential for noncomplementarity between the introduced mutations and surrounding regions of the protein. By this process we found that the W256Y mutation, representing a reversion back to the GlnRS sequence, improved $K_{\text{m}}[\text{tRNA}]$ and $K_{\text{m}}[\text{Glu}]$ each by two-fold compared to L1L2 GlnRS— probably by alleviating a steric clash with nearby Ile14 (Table 4; Figure 15D).[46] $K_{\text{m}}[\text{Glu}]$ in L1L2W256Y GlnRS is well below the intracellular Glu concentration in *E. coli* and is lower than measured for yeast GluRS and archaeal GluRSND (Table 4).[46, 102] The k_{cat} in L1L2W256Y GlnRS is identical to GluRSND, although it remains 15 to 30-fold lower than that of *E. coli* GlnRS or yeast GluRS. However, $K_{\text{m}}[\text{tRNA}]$ is 15 to 125-fold higher than measured in cognate aminoacylation by any of the three benchmark enzymes annotated in Table 1, despite the fact that no mutations at the tRNA^{Gln} interface were introduced. Therefore, while alterations to protein structure alone fully recreate a functional Glu binding site in GlnRS, the resulting hybrid enzyme remains deficient in pairing amino acid with tRNA. These observations underscore the

interdependence of amino acid and tRNA binding in GlnRS,[35, 38] as previously illustrated by the sharply elevated K_M [tRNA] for Glu-tRNA^{Gln} synthesis by the wt enzyme (Table 4) [37]. Although comparable protein engineering experiments have not been performed with other aaRS, it is likely that such substrate interdependence is embedded in the architectures of all aaRS from both structural classes.[1] This sets the stage for a fourth iteration of the rational engineering approach, but now considering improvements encompassing RNA nucleotides as well.

		k_{cat} (s ⁻¹)	K_M [tRNA] (μM)	k_{cat}/K_M [tRNA] (s ⁻¹ M ⁻¹)	K_M [GLU] (mM)	k_{cat}/K_M [GLU] (s ⁻¹ M ⁻¹)
<i>E. coli</i> GlnRS ^a	tRNA ^{GLN}	0.046 ± 0.013	19 ± 3	2.4 × 10 ⁻³	>750	9.5 × 10 ⁻⁴
C229R GlnRS ^b	tRNA ^{GLN}	3.2 ± 0.1 × 10 ⁻⁴	-	-	240 ± 10	1.3 × 10 ⁻³
S1 GlnRS ^b	tRNA ^{GLN}	5.0 ± 0.6 × 10 ⁻³	-	-	230 ± 17	2.2 × 10 ⁻²
L1L2 ^b	tRNA ^{GLN}	0.09 ± 0.02	7.6 ± 3.0	1.2 × 10 ⁻⁴	5.8 ± 0.5	15.5
L1L2 W256Y ^c	tRNA ^{GLN}	0.10 ± 0.04	4.8 ± 0.2	2.1 × 10 ⁻⁴	2.6 ± 1.0	38.5
L1L2 W256Y ^d	Q _E tRNA1	0.014 ± 0.01	1.2 ± 0.4	1.2 × 10 ⁻⁴	14 ± 7	1.0
L1L2 W256Y	Q _E tRNA2	0.05 ± 0.01	1.4 ± 0.7	3.6 × 10 ⁻⁴	53 ± 22	0.9
L1L2 W256Y	Q _E tRNA3	0.32 ± 0.05	2.4 ± 0.8	1.3 × 10 ⁻⁵	1.9 ± 0.2	168.4
L1L2 W256Y	Q _E tRNA4	0.069 ± 0.01	0.74 ± 0.07	9.3 × 10 ⁻⁴	3.3 ± 0.6	20.9
L1L2 R237D/R238E/W256Y	Q _E tRNA1	0.092 ± 0.01	3.4 ± 0.7	2.7 × 10 ⁻⁴	12 ± 5	7.7
L1L2 R237D/R238E/W256Y	Q _E tRNA4	0.29 ± 0.05	3.4 ± 0.5	8.5 × 10 ⁻⁴	3.0 ± 0.7	97.7
Benchmark enzymes						
<i>E. coli</i> GlnRS + GLN ^a	tRNA ^{GLN}	3.2 ± 0.5	0.31 ± 0.09	1.0 × 10 ⁻⁷	0.26 ± 0.04*	1.2 × 10 ⁻⁴
<i>M. thermotrophicus</i> GluRS ^{ND(c)}	tRNA ^{GLN}	0.12 ± 0.01	0.038 ± 0.010	3.2 × 10 ⁻⁶	6.2 ± 0.6	19.4
<i>S. cerevisiae</i> GluRS ^{D(e)}	tRNA ^{GLU}	1.6 ± 0.6	0.16 ± 0.03	1.0 × 10 ⁻⁷	39 ± 14	41.0

Table 4 -

^aReported in ref. 8. All values in this Table reflect glutamylation except for the benchmark represented by *E. coli* GlnRS for its cognate glutaminylation reaction (*E. coli* GlnRS + GLN).

^bReported in Rodriguez-.Hernández et. al. S1 GlnRS includes the following mutations: C229R/Q255I/S227A/F233Y.

S1L1L2 GlnRS adds the following mutations:

T214A/H215C/C216P/S218V/A220S/L221I/I224V/L231T/

V243I/L244I/D245E/N246A/I247L/T248G/P250R/ΔV251/H252K/R254Y/Y256W

^cReported in ref. Rodriguez-.Hernández et. al

^dSequences of hybrid Q/E tRNAs are provided in Figure 17.

^eReported in ref. Hadd et. al.

Results

These observations led us to consider a fourth rational design stage to explore the possibility that RNA structure is required for amino acid coding. Sequence comparisons among bacterial tRNA^{Gln} and eukaryotic tRNA^{Glu} species reveals conserved differences in the acceptor arm at positions 3-70 and 5-68, and in the globular core at positions 16 and 59 in the T and D-loops, respectively (Figures 17A, 17B).[29] We began by pairing GlnRS L1L2W256Y with a mutant tRNA^{Gln} in which the G3-C70 and G5-C68 pairs were altered to C3-G70 and U5-A68, respectively (Q/E_tRNA1; Figure 17). Remarkably, K_m [tRNA] is reduced four-fold by incorporation of these tRNA^{Glu} acceptor-stem base pairs, suggesting improved complementarity at the protein-RNA interface in a region close to the introduced eukaryotic GluRS peptides (Table 4). However, k_{cat} for Glu-tRNA^{Gln} synthesis is diminished seven-fold while K_m [Glu] is elevated by five-fold in this hybrid RNP, thus diminishing its overall catalytic performance.

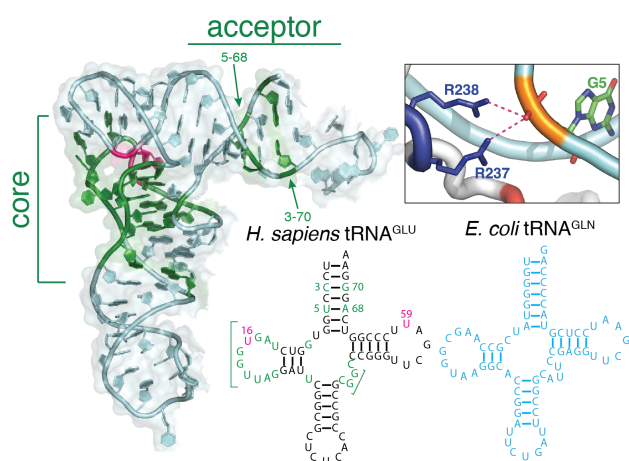


Figure 17 - tRNA structure and protein interactions. *Left*: tertiary structure of tRNA^{Gln} in the conformation bound to GlnRS. The 3-70 and 5-68 base pairs replaced in the acceptor stem are shown in green. The C16 and A59 core region nucleotides, replaced with U16 and U59 in some of the hybrid RNPs, are shown in red, while the remainder of the replaced tRNA core region is shown in green and consists of the C9G substitution together with replacements of the full D and variable loops. *Upper right*: Structural details of the protein-tRNA interface showing the backbone contacts made at the G5 phosphate of tRNA^{Gln} by Arg237 and A238 (dark blue). *Bottom right*: secondary structure depictions of human tRNA^{Glu} and *E. coli* tRNA^{Gln}.

These findings suggest an allosteric mechanism connecting protein-RNA contacts at base pairs 3-70 and 5-68 in the acceptor stem, with both the Glu binding site and the catalytic center where the two-step aminoacylation reaction occurs. To test this, we examined distinguishing bacterial GlnRS and eukaryotic GluRS interactions at these acceptor-stem positions. Arg237 and Arg238 in *E. coli* GlnRS, which interact with the sugar-phosphate backbone at nucleotide G5 (Figure 17), are replaced by Asp237 and Glu238 in GluRS. We therefore constructed the L1L2W256Y/R237D/R238E GlnRS enzyme and examined its capacity to glutamylate Q/E_tRNA^{Gln}. k_{cat} for Glu-tRNA^{Gln} synthesis is fully reconstituted in this RNP, and $K_m[\text{Glu}]$ is unchanged, but $K_m[\text{tRNA}]$ is elevated three-fold (Table 4). The reconstitution of k_{cat} indeed suggests intramolecular signaling of acceptor-stem backbone interactions to the catalytic center, but the other measurements indicate that this subregion of the protein-RNA interface does not operate as an independent module, but instead depends on other portions of the RNP for its function.

Next, we examined the role of the tRNA core domain. Introducing the C16U and A59U mutations into tRNA^{Gln}, to generate the Q/E_tRNA2 species for pairing with GlnRS L1L2W256Y, again improves $K_m[\text{tRNA}]$ but diminishes k_{cat} while sharply elevating $K_m[\text{Glu}]$ above 50 mM. This behavior is qualitatively similar to that observed with Q/E_tRNA1 (Table 4). However, the sizes of the D and variable loops differ between bacterial tRNA^{Gln} and eukaryotic tRNA^{Glu}, suggesting that C16U/A59U has disrupted tRNA structure. We therefore reconstituted the full human tRNA^{Glu} core region by introducing nine additional alterations in the D and variable loops to generate Q/E_tRNA3 (Figure 17). Kinetic analysis shows that L1L2W256Y GlnRS paired with

Q/E_tRNA3 is significantly improved for Glu coding compared to all prior RNP designs. Compared to L1L2W256Y paired with wt tRNA^{Gln}, k_{cat} for glutamylation is increased three-fold, $K_m[tRNA]$ is improved by two-fold, and $K_m[GLU]$ is maintained at about 2 mM (Table 1). These experiments demonstrate that the tRNA^{Glu} core region is an RNA determinant that specifies Glu for coding within the general architecture of class I aaRS RNPs.

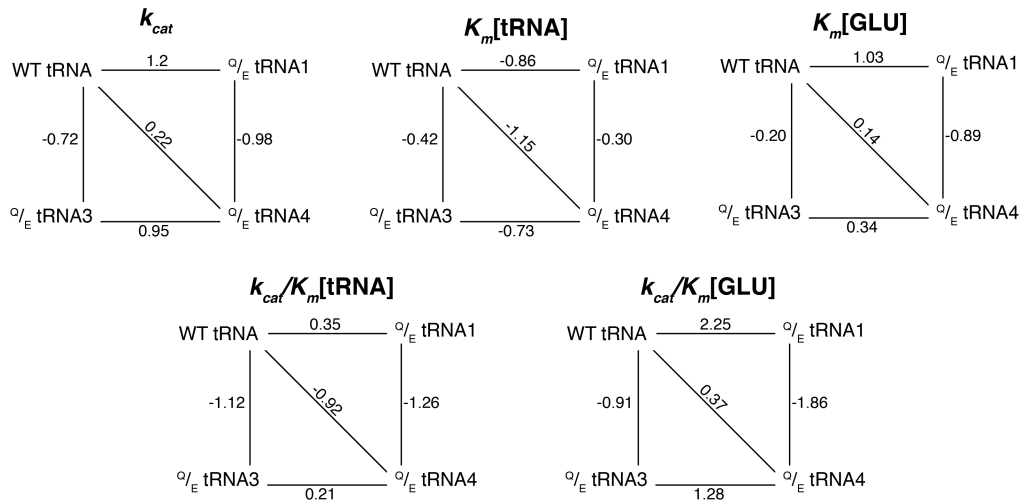


Figure 18 – Free energy component analysis at T= 310 K for the GlnRS S1L1L2 W256Y enzyme bound to Q/E_tRNA1 (altered at acceptor stem base pairs 3-70 and 5-68), Q/E_tRNA3 (altered in the tRNA core region), and Q/E_tRNA4 (comprising all the nucleotide replacements in Q/E_tRNA1 and Q/E_tRNA3). Free energy differences $\Delta\Delta G$ were computed from $(k_{cat})_{WT}/(k_{cat})_{MUTANT} = e^{(\Delta\Delta G/RT)}$, $(k_{cat}/K_m)_{WT}/(k_{cat}/K_m)_{MUTANT} = e^{(\Delta\Delta G/RT)}$, or $(K_m)_{WT}/(K_m)_{MUTANT} = e^{(-\Delta\Delta G/RT)}$.

Combining the acceptor stem and full tRNA core region replacements into one RNA (Q/E_tRNA4) has distinct effects on tRNA *versus* Glu-related kinetic parameters. $K_m[tRNA]$ shows substantial additivity between the two “single” substitutions in the acceptor stem and core (Q/E_tRNA1 and Q/E_tRNA3, respectively), and the “double” substitution in Q/E_tRNA4 (Figure 18). Therefore, enzyme-tRNA interactions at the inner elbow and acceptor stem regions appear to be largely independent of each other

with respect to a Michaelis parameter that approximates tRNA binding. Because of the additive effects, $K_m[\text{tRNA}]$ for the RNP composed of GlnRS L1L2W256Y paired with Q/E_tRNA4 falls below 1 μM and is elevated by just two-fold compared to the native GlnRS-tRNA^{Gln} interaction (Table 4). This demonstrates that rational design can be successfully applied to adjust this key steady-state kinetics parameter into a physiologically relevant range commensurate with intracellular tRNA concentrations.

In contrast, the GlnRS L1L2W256Y: Q/E_tRNA4 RNP exhibits little additivity with respect to $K_m[\text{Glu}]$; instead, the “double” substitution maintains a relatively low value for this parameter that primarily reflects the influence of the core region (Table 4; Figure 18). It is remarkable that the more distal tRNA^{Glu}-like core dominates the formation of the Glu binding site, since the introduction of positive RNA determinants in this region blocks the disruptive effect of the proximal acceptor stem base-pair swaps in Q/E_tRNA1. These data again emphasize that long-range allosteric signaling from the tRNA core region interface to the Glu binding site is a key aspect of Glu coding in this bacterial system.

Unlike all prior experiments based on subtractive mutagenesis, this rational engineering approach definitively demonstrates that RNA determinants are required for efficient amino acid-RNA pairing, at both k_{cat} and $K_m[\text{tRNA}]$ levels. Such RNA elements may represent vestiges of an early RNA world in which amino acid coding for protein synthesis was facilitated by direct amino acid-RNA interactions.[29] The exclusive role of protein in forming the amino acid binding pocket of contemporary aaRS RNPs obscures this function of the RNA, and indeed we have shown that tRNA^{Glu} nucleotides are not required to optimize $K_m[\text{Glu}]$ in the engineered enzymes. However, the protein

binding site for Glu does not function optimally in RNA pairing unless RNA nucleotides that encode Glu are also included in the RNP. We suggest that optimization of allosteric linkages operating between the mutated tRNA core region and acceptor stem base-pairs, by rational design of the intervening protein elements (Figure 3), should further improve function. $k_{\text{cat}}/K_{\text{m}}[\text{tRNA}]$ for the engineered RNPs remains 30-100 fold below that of the naturally occurring benchmark enzymes (Table 4), and improving this parameter likely requires repair of protein structure in peptides adjacent to the R237D/R238E substitutions (Figure 17).

Discussion

Our findings provide insights that should be useful in creating new aaRS that encode ncAAs for synthetic biology applications (Figure 19). Randomizing the local amino acid binding sites of archaeal TyrRS and PylRS generates enzymes capable of ncAA incorporation, an outcome that likely rests on the fact that the aaRS scaffolds were not under selective pressure to resist these ncAAs. However, our experiments suggest that directed evolution will not succeed in generating highly efficient enzymes unless the second/third shell protein residues and the allostery between amino acid and tRNA substrates are also considered. All aaRS are large multidomain proteins that bind tRNA in a mutual induced fit process that likely proceeds by formation of an initial encounter complex, followed by first-order rearrangements that may also be affected by the binding of ATP and amino acid.[1] These common features observed in X-ray structures predict that allosteric functional linkages between tRNA and amino acid binding, and the

broad dispersal of amino acid specificity determinants, will also be prevalent if not universal among the aaRS.

Although the design of ncAA-coding aaRS cannot benefit from structure/sequence databases, one approach that may allow generation of more catalytically proficient enzymes could be to rationally engineer sets of hybrid methanogen TyrRS and PylRS that incorporate distal structural elements from other homologous class 1c and class 2c aaRS, respectively. This would substantially expand the number of subtly distinct scaffolds for use in genetic selections (Figure 19). The exquisite sensitivities of naturally occurring aaRS to mutational perturbation suggests that even relatively small differences in active site structures and dynamics, within families of such rationally engineered enzymes, could generate significant differences in capacities for ncAA

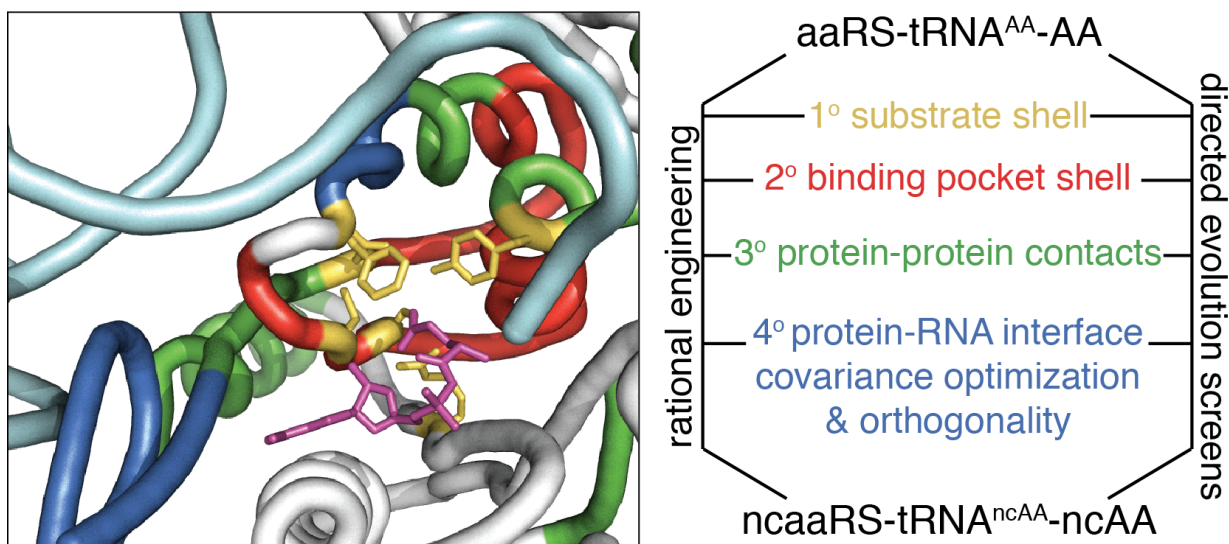


Figure 19 - Strategy for rational design of amino acid specificity in the aaRS, illustrated in *E. coli* GlnRS: (i) the S1 site is depicted in yellow; (ii) the distal L1 and L2 loops are depicted in red; (iii) surrounding regions of the protein structure that contact S1, L1 and L2 are depicted in green; (iv) peptides that directly contact tRNA in regions that assist amino acid-RNA pairing specificity are depicted in dark blue. Binding of enzyme at the tRNA inner elbow is shown at the bottom left; the separate contacts at the tRNA acceptor stem are shown at the top center. The tRNA backbone is shown in light blue and the glutamyl adenylate analog is shown at bottom center in magenta.

incorporation when subjected to directed evolution. Systematic rational engineering of methanogen TyrRS and PylRS along the lines described here would allow assessment of the nature and spatial distribution of amino acid specificity determinants, and the extent to which they may reside in the RNA. Precise choices of residues to be randomized, within the context of rationally engineered scaffolds, will clearly best be made when a quantitative database of structure-function information is available.

A second insight arising from these experiments is that directed evolution strategies that include selections for optimizing orthogonality in the tRNA should account for the possibility that the derived tRNAs may acquire nucleotides that render them less efficient for incorporation of a desired ncAA. This notion emerges from our finding that tRNA^{Glu} core elements are positive determinants for encoding Glu. Careful measurements of kinetic parameters for aminoacylation of ncAAs by selected RNPs are required to evaluate this and other detailed properties of ncAA-encoding aaRS.[103] RNA and protein sequences in ncAA-encoding RNPs should be optimized for both orthogonality and efficiency of aminoacylation.

Finally, it is worth noting that none of the efficiently functioning hybrid RNPs constructed in our studies alter any part of the tRNA anticodon arm, despite the important role for tRNA anticodon loop nucleotides in specifying tRNA^{Glu} and tRNA^{Gln} identities.[35, 38] Although further experiments are required to fully evaluate this notion, it appears that anticodon recognition by GlnRS may be less fully integrated into the protein-RNA allosteric network than we have previously suggested.[35] Perhaps either GlnRS-like or GluRS-like anticodon recognition may be substantially compatible with

coding of either amino acid. The relative independence of anticodon recognition is beneficial to the further development of orthogonal translation systems.

Chapter III

*Crystallography of *Saccharomyces cerevisiae**

Arc1p-GluRS-tRNA^{Glu} ternary complex

Introduction

As discussed in the previous chapters, a ubiquitous trait in eukaryotic translation is the existence of multi aminoacyl-tRNA synthetase (MARS) complexes (Figure 6). Complex formation is typically generated due to the domain extensions, also associated with eukaryotic aaRSs [16]. Furthermore as described previously and in Chapter I, it had been demonstrated that these extensions and additional RNA binding proteins facilitate improved aminoacylation activities [29]. While until recently we have begun to understand the functional characterization of eukaryotic aaRSs and the functional role of appended domains, minimal structural characterization exists.

Described in Chapter I, *S. cerevisiae* Arc1p forms a complex with GluRS and MetRS generating a small eukaryotic MARS complex. Arc1p interacts with these two aaRSs via two independent N-terminal GST domains on either enzyme and has been shown to form a tight binding complex ($K_d \sim 50$ nM) (Figure 19). A key functional effect of additional RNA binding domain increases GluRS activities by enhancing tRNA binding and modulating amino acid affinity. Previous studies had also demonstrated similar behaviors for MetRS, as mentioned [104]. While the *S. cerevisiae* Arc1p-GluRS-MetRS complex has now been biochemically well studied, the only structural information that exists are binary complexes of the N-terminal domains of GluRS-Arc1p and Arc1p-

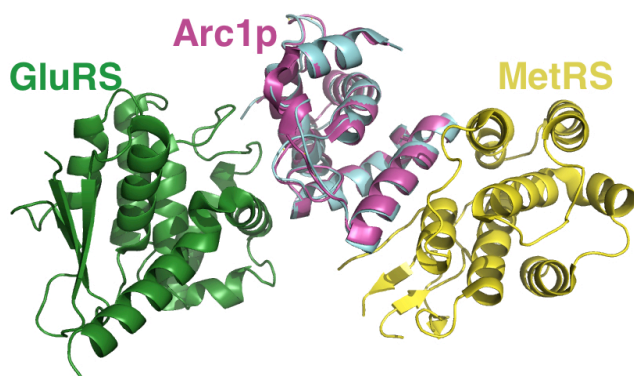


Figure 20 – Structural overlay of the interacting domains of GluRS (Green) and MetRS (Yellow) with Arc1p. Two separate structures were solved between GluRS with Arc1p and MetRS with Arc1p. The model was generated aligning Arc1p together creating an RMSD = 0.67 Å.

MetRS; there is no structural characterization of full length *S. cerevisiae* GluRS, Arc1p or MetRS bound or unbound nor with their respective tRNA substrates (Figure 13, 20) [83, 105].

Given our findings, we sought to solve the X-ray crystal structure of GluRS bound with Arc1p in complex with tRNA^{Glu} and glutamate-adenylate active site inhibitor analog (ESI). This structure would be the first eukaryotic aaRS bound to an auxiliary protein, with or without tRNA, as well as the first eukaryotic β -GluRS structure. Furthermore we felt we could successfully achieve stable crystals given the relatively less complex quaternary structure with only two tight binding proteins ($K_d = 50$ nM) and a tRNA with high binding affinity ($K_d = 160$ nM) [74]. This is in comparison to the *H. sapiens* MARS complex that contains 11 separate proteins. Our proposed ternary complex structure would provide key insights into protein-protein complex formation between GluRS and Arc1p. Moreover, it would shed light on how each of these proteins interacts with the tRNA^{Glu} and glutamate substrates.

S. cerevisiae GluRS has high specificity for tRNA^{Glu} that is further enhanced by Arc1p. When considering the phylogenetic distribution of GluRSs and GlnRSs (Figure 9), we currently have structures of bacterial α -GluRSND and α -GluRS^D, archaeal β -GluRSND and bacterial and eukaryotic β -GlnRS, thus a key missing structure is a

eukaryotic β -GluRS^D, further justifying these experiments. Shown in figure 13 is a structural overlay of a β -class GluRSND and GlnRS, modeled with tRNA^{Gln}. This comparison showed three main insertions in GlnRS that gave rise to tRNA^{Gln} as compared to GluRSND. These insertions include the N-terminal domain in GlnRS and two protruding loops that interact with the tRNA 1-72 pair and 36 positions; all three nucleotides are found to be identity elements in these systems. Solving the *S. cerevisiae* β -GluRS^D would provide further insight to what protein structural elements co-evolved to gain tRNA^{Glu} specificity; this could be especially informative comparing both GlnRS-tRNA^{Gln} and the GluRSND structures. Furthermore, solving a structure with Arc1p bound would shed light onto what elements give rise to the ten-fold improved tRNA binding.

From a protein engineering standpoint and building on the results from Chapter II, solving a β -GluRS^D structural could improve mutational design. As mentioned, the only GluRS structures available are of α -class and β -GluRSND, however our engineering efforts in Chapter II were seeking to change the specificity of a β -GlnRS towards a β -GluRS^D. This structure would help us understand potential differences within the glutamate-binding pocket in a β -class discriminating enzyme relative to other GluRS enzymes thus improving our mutational design for synthetic biology applications.

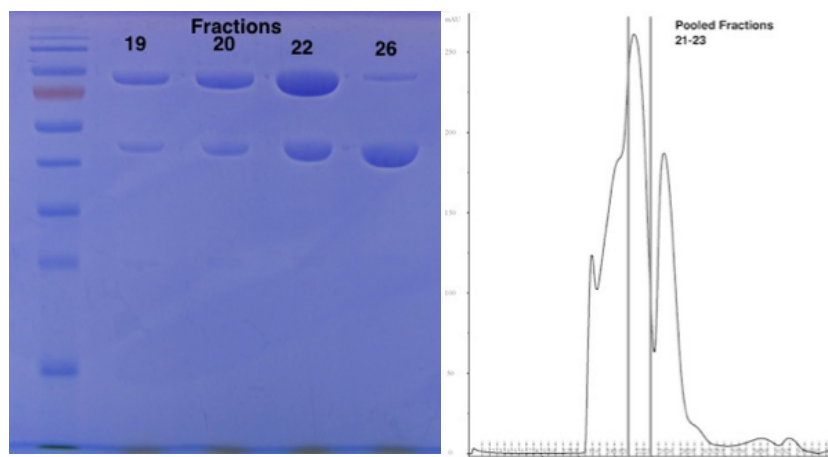
The fundamental underpinnings that solving this structure would elucidate justified the invested time finding crystallization conditions. Initial high-throughput screens found a single lead. Unfortunately after exhaustive efforts, I was unable to improve conditions that would allow us to move the project towards X-ray analysis. This project does however provide conditions for sample preparation and a baseline for crystallization.

Results – Protocols

Protein-RNA purification – See protein and tRNA purification.

FPLC purification

After purification of the individual GluRS and Arc1p proteins, do not move either into glycerol storage conditions, but leave at high sodium chloride buffer conditions and store at 4°C. To prepare for the purification, rinse and equilibrate the FPLC system attached with a GE Healthcare Superdex 200 10/300 preparatory gel filtration column in buffer containing filter sterilized 50 mM Tris-HCl pH 7.5, 500 mM sodium chloride, 10 mM β -mercaptoethanol. The column should be rinsed with at least five column volumes of buffer before preparing the protein sample. To prepare the protein sample, mix a two-fold excess of Arc1p relative to GluRS with a total protein concentration of 10 mg/mL at final volume ~1 mL. This ensures to maximize GluRS bound to complex and to easily be able to resolve the two peaks: one containing GluRS-Arc1p and the other excess unbound Arc1p. Run sample at 0.4 mL/min for two column columns and collect the complex peak (elution time ~8 minutes). Dialyze protein complex into 50 mM Tris-HCl (pH 7.5), 10 mM NaCl and 20 BME and concentrate down to 10 mg/mL. If the crystallography experiment was conducted the next day, the protein would be stored on wet ice set at 4°C.



Gel image (Left) of the fractions run from the FPLC purification (Right). Lane 25 from the gel shows the excess of Arc1p of the purification.

Complex Preparation

Use freshly purified protein complex when preparing the GluRS-Arc1p-tRNA-ESI complex. The protein-tRNA complex was first created to generate a 1:1.2 protein:tRNA ratio, with a final protein concentration of 8 mg/mL. The tRNA was first refolded by heating the tRNA stock and 50 mM Tris-HCl (pH 7.5) 10 mM MgCl₂ separately for three minutes and slow cooled to ambient temperature in a heat block. After mixing protein and RNA together, ESI was added to a final concentration of 1 mM with a final volume to set up both high-throughput and manual screens is 30 μ L. The complex was then incubated on ice for at least one hour before running the experiment.

Tested Screens and Ideas

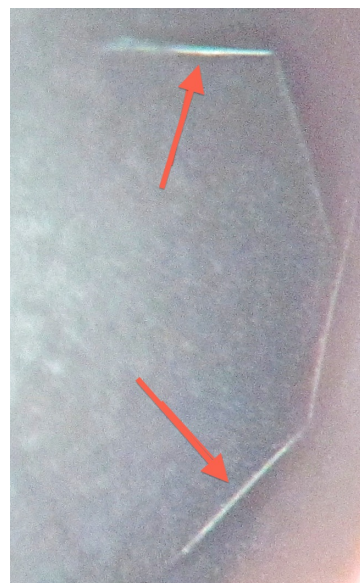
Using the high-throughput liquid handling robot, the following screens were used using 150 nL drop mixed with 150 nL well solution (300 nL total). All screens tested were incubated at both 4°C and 15°C.

Hampton High-Throughput Screens:

- Crystal Screen HT
- Index HT
- Natrix HT

PegRX HT, Peg/Ion HT, and SaltRX HT conditions were each tested within the previously mentioned tested screens and did not feel it was necessary to further probe these “fine-tuned” screens.

A single hit was found and confirmed using UV imaging. The buffer conditions were 0.2 M ammonium acetate, 0.1 M BIS-Tris (pH 6.5), and 45% (v/v) (+/-) 2-methyl-2,4-pentandiol (MPD). Small spindle needles were observed after three days of incubation at 4°C. The crystals were however covered in aggregate making it very difficult to work with and clean. These crystals were not able to repeat after scaling up using manual pipetting screens. In order to improve crystal growth, I attempted two other experiments - microseeding using the same conditions and also a divalent additive screen using the same mother liquor conditions at



Crystal image of Arc1p-GluRS-tRNA^{Glu} and ESI inhibitor grown after three days at 4°C

4°C. The microseeds provided no further growth. The following divalent salts were tested at both final 1 and 10 mM concentrations: MgCl₂, Mg(OAc)₂, MnCl₂, CoCl₂, FeCl₂, NiCl₂, CaCl₂, SrCl₂, BaCl₂, and CdCl₂.

Future experiments need to be conducted to reproduce the initial hit continuously. From here improving the complex purification should be further refined; changing the FPLC conditions and also purification of a ternary GluRS-Arc1p-tRNA complex are two places to start. Also attempting to vary the complex concentration within the drop.

Chapter IV

Concluding Remarks

Aminoacyl-tRNA synthetases fulfill a unique role in biology and are key enzymes for maintaining the genetic code. A theme throughout this work has been refining the concept of the genetic code, which more broadly interpreted is deciphering nucleic acid based information into amino acid products. Based on this, aaRS function is fundamental in translating the code as they *read* the RNA based message and ensure the ligation of amino acid to the cognate tRNA. This functional pairing between the tRNA and the amino acid requires an extraordinarily level of macromolecular specificity and moreover, this task maintains the fidelity of the code. This work has sought to understand the mechanisms of specificity within the glutamine and glutamate coding defined by the RNA coding signatures within the eukaryotic lineage. Using *S. cerevisiae* GluRS and GlnRS as model systems, this work found the glutamate and glutamine tRNA identity elements in eukaryotes. These tRNA elements are readout directly and indirectly by their cognate GlxRS. Furthermore, these tRNA genetic signatures were found to be at mutually exclusive positions relative to tRNA^{Glu} and tRNA^{Gln}. This trend accomplishes two things: allows the tRNA signature to be positive determinants for the cognate GlxRS and anti-determinant for the non-cognate GlxRS. Therefore the tRNA elements ensure the proper selection and synthesis of aminoacyl-tRNA. Evolutionarily speaking, this is of utmost importance given the phylogenetically relatedness between GluRS-tRNA^{Glu} and GlnRS-tRNA^{Gln}, each evolving from a common progenitor GluRSND.

Therefore having distinct RNA elements at the same positions can ensure complete rejection between noncognate GlxRS-tRNA pairs.

While seeking to understand how these RNA elements are readout by their GlxRS, this work additionally discovered a novel function the tRNA fulfills in amino acid selection within the aaRS. Previous studies have shown, which is further corroborated in the *S.cerevisiae* systems, the existence of binding interdependence between tRNA and amino acid substrates. In Chapter I, mutating the tRNA^{Glu} and tRNA^{Gln} core identity elements function modestly affects tRNA specificity and makes contributions to amino acid specificity. This was demonstrated by an improved k_{cat}/K_m [amino acid] upon the addition of core tRNA identity elements into a noncognate tRNA substrate. For example, K_m [Gln] improved two-fold for *S. cerevisiae* GlnRS after adding the 15,16,48 and 59 nucleotides from tRNA^{Gln} into noncognate tRNA^{Glu}. The same overall trend held for glutamate specificity for *S. cerevisiae* GluRS.

This led to the hypothesis that tRNA identity elements may serve to not only function towards tRNA specificity but also assist in aaRS specificity for the cognate coded amino acid. Previous lab members developed an enzymatic system to test this hypothesis using an engineered aaRS. Using a rational engineering approach, previous studies found the minimal protein substitutions to fundamentally change the amino acid specificity in *E. coli* GlnRS from glutamine and glutamate. The engineering design deliberately ensured all mutations made did not contact along the protein-tRNA interface. Yet while successfully reconstituting glutamate specificity, the catalytic rate constant and tRNA specificity were still orders of magnitude lower than that of naturally occurring aaRS systems. Applying the findings from Chapter I, the work in Chapter II

proved this hypothesis that tRNA identity elements are capable of modulating the specificity for the corresponding amino acid. Incorporating tRNA^{Glu} identity elements improved the activities of our engineered system, specifically a four- and five-fold improved glutamate and tRNA specificity, based on k_{cat}/K_m respectively.

Moreover this novel finding demonstrates a key function translating the genetic code. It shows that aaRSs act to bridge the stereochemical identities of amino acid and tRNA together. Yet aaRSs utilize RNA identities to assist in the selection of the coded amino acid. Thus poor-productive tRNA binding will also weaken amino acid specificity as well. These mechanisms assist in genetic code fidelity ensuring the proper coupling of amino acid and tRNA.

Aminoacyl-tRNA synthetases in the future

Decades of research provide a foundation for continued studies on aaRSs. These include drawing on efforts to understand mitochondrial aaRSs, connecting the translation apparatus to cellular metabolism and using aaRSs for synthetic biology purposes.

Drawing back to the story from Chapter I are continued questions with Arc1p and its effects on GluRS activity. Hinted throughout this thesis is *S. cerevisiae* GluRSs role in both cytoplasmic and mitochondrial translation. GluRS primarily is involved in cytoplasmic translation where it efficiently aminoacylates cytoplasmic tRNA^{Glu} with the assistance of Arc1p. Mitochondrial protein synthesis is carried out by a complete set of translation genes. These include a complete set of mitochondrial specific tRNAs and

ribosomal RNAs, all transcribed in the mitochondria as well as a complete set of aaRS. However, mitochondrial aaRSs are nuclear encoded but have an N-terminal mitochondrial targeting sequence. Given the bacterial origin of the mitochondria, these aaRSs are more similar to bacterial than eukaryotic; this also includes the use of the two-step pathway for Gln-tRNA^{Gln} synthesis using the amidotransferase (AdT) (Figure 8). As with most eukaryotes, *S. cerevisiae* encodes a mitochondrial Gln-tRNA AdT, however the mitochondrial GluRS is discriminating, leaving the question, how is Glu-tRNA^{Gln} synthesized in the mitochondria, the substrate for the AdT [76, 105]? It had been demonstrated that the cytoplasmic GluRS fulfills this missing activity, where it misacylates the mitochondrial tRNA^{Gln} with glutamate. Furthermore, the cytoplasmic sub-cellular location is modulated by sequestration due to Arc1p: cytoplasmic GluRS is maintained in the cytoplasm when complexed to Arc1p but is transported to the mitochondria when out of complex. Therefore, Arc1p modulates cytoplasmic GluRS specificity; making solving the proposed Arc1p-GluRS crystal structure discussed in Chapter III intriguing.

We sought to biochemically characterize cytoplasmic GluRSs ability to misacylate mitochondrial tRNA^{Gln} by obtaining *in vivo* tRNA^{Glu} and tRNA^{Gln} from both the cytoplasm and mitochondrial isolates using Chaplet Column Chromatography (described in the methods). Unlike the studies done in Chapter I that used *in vitro* transcript tRNAs, studying the mitochondrial tRNAs required the use of the *in vivo* tRNAs because both glutamate and glutamine isoacceptors encodes a post-transcriptional modified nucleotide at position 34 that is required for aminoacylation. Also described in the methods in more detail, I was unable to obtain mitochondrial tRNAs at either high

enough concentrations or with the ability to be substrates for the nucleotidyltransferase enzyme, used to prepare ^{32}P radiolabeled substrates for the kinetics assay.

Another interesting aspect of the *S. cerevisiae* Arc1p and GluRS system is what regulates GluRS sub-cellular location. Yeast cells typically metabolize sugars under fermentation for ATP production, minimizing oxidative phosphorylation activity in the mitochondria. When sugars are scarce, the cellular metabolism shifts to respiration, therefore higher energy output by the mitochondria [76, 105]. This increase in mitochondrial activity requires the mitochondrial translational apparatus to increase, thus an increase in Glu-tRNA^{Gln} synthesis. This demand therefore would require the cytoplasmic GluRS in both sub-cellular locations, some bound to Arc1p in the cytoplasm and some unbound and transported to the mitochondria. It has not been clearly demonstrated what causes a higher fraction of GluRS to be transported to the mitochondria, however one hypothesis is carbon source regulation (either at the transcriptional or translational level) of the Arc1p gene [105]. Perhaps under high carbon source, Arc1p gene output is higher but is down regulated under a poor/low carbon source. Nonetheless a clear connection can be drawn between GluRS and Arc1p in relation to carbon metabolism, thus creating a crossroads between energy metabolism and translation. Interestingly Arc1p homolog in humans (AIMP1) has also been shown to be involved in carbon metabolism [16, 106-108]. Studies in mice showed glucose starvation induces a spike in AIMP1 concentration in blood plasma. In these studies, Park et. al. could induce gluconeogenesis in mice pancreatic cells by intravenous infusion of AIMP1 [106-108]. Taken together, it shows an interesting

parallel that both translation factors (Arc1p and AIMP1) involvement in carbon energy metabolism.

Further studies to understand how these translation components regulate energy metabolism would be worthwhile and could shed light on potential disease pathways, such as diabetes. If it could be shown that Arc1p and AIMP1 both regulate similar carbon metabolism pathways, this could allow *S. cerevisiae* to be used as a potential simpler model system whose findings could be later translated into a more clinically relevant system (e.g. mammalian pancreatic cell lines).

Continued studies on eukaryotic aaRSs, specifically the mitochondrial aaRSs would not only further our understanding in aaRS involvement in energy metabolism as discussed, but also could enhance the development in genetic code expansion. Very little is understood about mitochondrial aaRSs thus a general understanding on their activities would expand our basic science knowledge. Furthermore, mitochondrial tRNAs have been shown to be substantially unique relative to all other tRNA systems [81]. This uniqueness is due to many mitochondrial tRNAs lack some of the highly conserved tRNA elements, as far as some missing entire D-arms (Figure 2). However, even with these missing elements, these tRNAs are still functional, thus their corresponding aaRSs have coevolved mechanisms to compensate for these changes, which are unknown. As discussed in Chapter II, one of the major limiting factors in genetic code expansion is reengineering efficient aaRS-tRNA systems that are also orthogonal. I propose the use of eukaryotic mitochondrial aaRS-tRNA system for genetic code expansion scaffolds given the likelihood these aaRS-tRNA pairs would be orthogonal to most host systems due to the fundamental differences in the tRNAs. It

would be extraordinarily surprising that importing a mitochondrial tRNA with a missing D-arm to have cross-compatibility hurdles needed to overcome.

Most genetic code expansion efforts have found success using the archael TyrRS and PylRS systems due to their engineerability and their orthogonal tRNAs. Expanding on potential engineering scaffolds for synthetic biology efforts would be useful. One proposal would be to utilize a mitochondrial TyrRS-tRNA^{Tyr} system for engineering since TyrRS has already been found useful for unnatural amino acid incorporation and the amino acid binding pocket has proved to be capable of engineering. Coupling this concept of mitochondrial aaRS with the engineering strategy outlined in Chapter II could generate new research programs in the growing field in synthetic biology.

Methods

tRNA Preparation

In vitro tRNA transcription

All *in vitro* RNA transcription began by generating a DNA template following previously defined protocols [109]. DNA templates were synthesized using Klenow DNA polymerase fragment and two DNA oligos that contained a 10-base complimentary overlap region. The template strand contained a T7 RNA polymerase promoter (5'AATTCCTGCAGTAATACGACTCACTATA_gene). The first two bases of non-template strand contained 2'O methyl ribose sugar modifications. This assists in efficient RNA polymerase ejection and creating clean transcription product 3' ends. The oligonucleotides were ordered from IDT at 100 nmol using standard desalting purification. DNA templates were prepared using 4 μ M of each primer and 50 units of Klenow fragment DNA polymerase in 500 μ L total volume. After extension, the DNA was purified by ethanol precipitation and resuspended in 400 μ L of TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA).

Two mL *in vitro* transcription reactions were then setup using 200 μ L of the prepared DNA template. Transcription reactions were conducted overnight at 37°C in 40 mM Tris-HCl pH 8.0, 25 mM magnesium chloride, 2 μ M spermadine, 1.5% Triton-X-100, 40 mM dithiothritol, 4 mM of each NTP (pH 7.0), 10 μ g pyrophosphatase and 200 μ L 0.4 mg/mL T7 RNA polymerase. RNA was separated from protein by two phenol-chloroform (pH

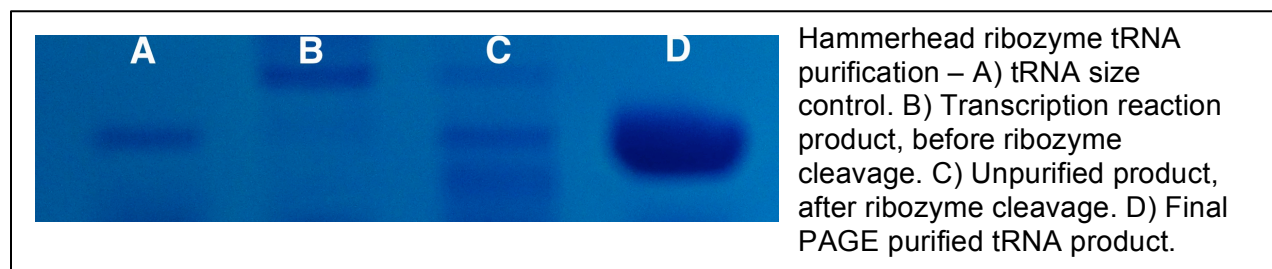
7.5) extractions and then concentrated to 300 μ L using a 10k MWCO Amicon filter. RNA was further purified on a 10% 8M urea denaturing TBE-PAGE gel. After four hours of electrophoresis at 50 watts, RNA bands are observed by UV shadow and excised from the gel. RNA was extracted from gel pieces by soaking into TE buffer overnight at room temperature. Final purified tRNA product was then concentrated to 200 μ L and buffer exchanged using fresh TE buffer. Under these conditions, typical tRNAs with a G1 yields ~2-3 milligrams of purified material. Non-G1 containing tRNAs yield ~0.5 mg at this scale.

Hammerhead ribozyme RNA transcription

If the RNA product does not contain a G1, at times this may cause for inefficient transcription using T7 RNA polymerase and can be alleviated by the addition of a self-cleaving hammerhead ribozyme 5' of the gene, after the promoter [110]. The following sequence would be added 5' to the gene (this contains the T7 promoter and ribozyme):
5'AATTCCTGCAGTAATACGACTCACTATAGGGGAGACGGACTGATGAGTCCGTGAG
GACGAAACGGTACCCGGTACCGTC_gene. (underlined portion is hammerhead).

The DNA templates would be generated in two phases. Given the template length, four DNA primers are required and designed the same way as before for normal transcription reactions using 10 basepair overlap. For tRNA genes, these templates are ~150 basepairs in length. However, these Klenow products generate numerous by-products. The template of interest can be purified by PCR amplification using primers complimentary to the 5'end of T7 promoter and 3' of the tRNA gene. After PCR, the

transcription reaction is carried out under identical conditions. After transcription but before phenol-chloroform extraction, the ribozyme must be cleaved. The cleavage reaction can be carried out to 80% efficiency by diluting the transcription reaction five-fold using the same transcription buffer and incubated at 60°C for four hours. From here the tRNA is purified as before and typical yields are near 0.5 mg at this scale.



³²P labeling tRNA for enzyme kinetics experiments

All kinetics experiments contained ³²P tRNA prepared following an adaptation of the Wolfson-Uhlenbeck protocol [111]. Eighty microliter labeling reactions were set up by refolding 2 μM of tRNA into CCA reaction buffer. The final concentrations of the reaction buffer is 25 mM HEPES pH 7.4, 6.25 mM magnesium chloride, 0.32 mM dithiothritol, but stored at 4x. After refolding, 2 μL of 100 mM pyrophosphate and 2.5 μL of ³²P α-ATP were added to the reaction. The reaction was then pre-incubated at 37°C for two minutes and then 6 μL of CCA enzyme was added. The reaction was then incubated at 37°C for one minute and quickly spun down to remove condensation and avoid radioactive contamination. One microliter of 1 mg/mL pyrophosphatase is added to the reaction and incubated at 37°C for five minutes. The reaction is quenched and protein removed by two 100 μL of phenol-chloroform (pH 7.5) extractions. The extracted

labeled RNA was then purified on a non-denaturing 10% TBE-PAGE gel. The gel was run at a constant 350 volts for 45 minutes. Using a small amount of RNA loading dye doped with ^{32}P was spotted on to the gel after being taken down and glass plates removed. This is to help guide where to excise the tRNA band from the gel. The gel is then exposed to a phosphor screen for 20 seconds and scanned on the phosphorimager. Place a printed copy of the scan without reducing the size (1:1) under gel using the dyed/radioactive spots as guides. Using a sterilized razor cut the gel band containing the tRNA and place it into a 1.5 mL eppendorf tube. Flash freeze the gel piece using dry ice. After frozen, soak in TE buffer overnight at 4°C. The next day, separate the TE buffer, now containing the labeled tRNA, into a fresh tube and use accordingly.

In vivo tRNA isolation from S. cerevisiae or E. coli

Protocols for *in vivo* tRNA was adapted from established protocols [5, 112]. Isolating *in vivo* tRNA follows the same protocol for both *E. coli* and *S. cerevisiae*, except for the obvious cell culture steps. One liter of *E. coli* cells are grown to mid log phase ($\text{OD}_{600} \sim 0.5$) in LB media at 37°C.

S. cerevisiae cultures begin by starting a five milliliter overnight culture in YP-2% dextrose (YPD) media shaking at 30°C. Overnights are then transferred to 20 mL of YPD for another hour at 30°C before transferring to 1 L of YPD. If attempting to recover tRNA from mitochondria, instead of growing in 1 L dextrose containing media, use 2% galactose as the carbon source. Galactose promotes respirative metabolism and the

ability to reach higher OD unlike glycerol, which will force respiration instead of fermentation but the growth rate is much slower. As with *E. coli*, grow cultures to mid log phase and then harvest by centrifugation.

Preparing cell lysates for both *E. coli* and *S. cerevisiae* cultures are identical. Cell pellets are resuspended in 8 mL of buffer containing 20 mM Tris-HCl (pH 7.5) and 20 mM magnesium acetate. Cells are lysed by the addition of 10 mL acid phenol-chloroform, which is then heated at 65°C for 10 minutes. After heating, the lysate is transferred to room temperature and shook for another hour. Bulk RNA is the extracted within the aqueous layer after centrifugation. Bulk tRNA is isolated from ribosomal RNA and mRNA by purification on a 100 mL 8 M urea-TBE PAGE gel run for four hours at 25 watts. The tRNA band is visualized by UV shadow, excised and eluted in 10 mL of TE buffer. The buffer is concentrated and buffer exchanged into 400 μ L 30 mM Tris-HCl pH 7.5, 1.0 M tetraethylammonium chloride, and 15 mM ethylenediaminetetraacetic acid.

The individual tRNA isoacceptors are purified from bulk tRNA using a DNA hybridization, pull-down method, called Chaplet Column Chromatography [5, 112]. Purification columns are created in empty columns fit into 1.7 mL eppendorfs using streptavidin sepharose beads. The probes are created using a 5'biotinylated 30-35 nucleotide DNA oligo; the oligo is designed to hybridize to a distinct portion of the tRNA, (I have had best luck creating a 35-mer oligo that spans acceptor arm, D-arm and anticodon, and avoiding the T-arm with a $T_M \sim 70^\circ\text{C}$). The column is prepared by combining biotinylated DNA and streptavidin beads in 450 μ L of 10 mM Tris-HCl (pH 7.5), 400 mM sodium chloride and 5 mM EDTA and then incubated at room temperature for 20 minutes. Unbound oligo is removed by centrifugation and the column is washed

with TE buffer. After washing, the resin is then equilibrated into the same bulk tRNA buffer, 30 mM Tris-HCl pH 7.5, 1.0 M tetraethylammonium chloride, and 15 mM ethylenediaminetetraacetic acid.

The bulk tRNA is added to the purification column and heated at 65°C for 10 minutes in a dry bath heat block. The heat block is removed and slow cooled allowing the tRNA isoacceptor to anneal to the DNA probe. Unbound tRNA is removed by centrifugation. The resin is then washed with five-column volumes of room temperature TE buffer. Measure A260 of each sample using the NanoDrop to monitor the progressive washing steps ensuring a zeroed baseline. After the A260 baselines, two more subsequent washes are performed using TE buffer but at 40°C. This ensures removal of non-specific binding, and confirmed by A260 measurements. The target tRNA is then eluted from the column by resuspending the resin with 400 µL of TE buffer and incubated at 80°C for three minutes, which is then eluted by centrifugation. The elution is then concentrated to a desired volume (typically ~40 µL) and purity assessed by PAGE. Typical purifications typically yield 50 µg of tRNA from 1 liter of cells.

S. cerevisiae mitochondrial tRNA note

As previously noted if preparing mitochondrial tRNA from *S. cerevisiae*, grow cell cultures using galactose as the carbon source instead of dextrose. Cell lysates are prepared as previously mentioned since using phenol-chloroform and heat will lyse both the fungal cell wall as well as the mitochondrial membranes. However a major technical difficulty preparing mitochondrial tRNA is the overall yield. Even scaling culture growths

up to four liters, typical yields for *S. cerevisiae* mitochondrial tRNA^{Glu} and tRNA^{Glu} were only about one microgram to hundreds of nanograms. Furthermore these tRNAs were poor substrates for the nucleotidyltransferase enzyme used for CCA labeling in the kinetics assay. The poor yield and inability to efficiently label the tRNA did not allow us to characterize these tRNAs. Furthermore, using *in vitro* transcripts was not a viable option for this characterization as both glutamate and glutamine isoacceptors coded for a U34 that is post-transcriptionally hypermodified with sulfur the 2-position and methylaminomethyl at the 5-position, both required for efficient aminoacylation activities.

Protein Purification

S. cerevisiae GluRS and Arc1p Purification

Both GluRS and Arc1p and purified identically and each gene is within the *E. coli* expression vector pET28(b) between NcoI and XhoI sites. This allows for each to contain a non-removable C-terminal His6 tag. Each construct was expressed in Novagen Rosetta2 DE3 pLysS under kanamycin and chloramphenicol selection. Five-milliliter overnight cultures were grown to saturation, which were used to inoculate two 1-liter cultures of LB/Kan/CAM media. Cultures were then induced with 0.5 mM IPTG after reaching OD₆₀₀ 0.7. After four hours of induction at room temperature, cultures were harvested by centrifugation at 4000 rpm for 15 minutes. Cell pellets were stored at -20°C.

Pellets were resuspended in 20 mL of buffer containing 50 mM Tris-HCl pH 7.5, 500 mM sodium chloride, 5 mM β -mercaptoethanol, 5 mM imidazole, Roche protease inhibitor tablet (no EDTA), 20 mg lysozyme and 200 μ g of DNaseI. Cells were then lysed by sonication and clarified by centrifugation for 45 minutes at 15,000 rpms. Clarified lysates were then filtered through a 0.45 μ m syringe filter and then allowed to incubate with Ni-NTA resin for one hour at 4°C in a 50 mL falcon tube. The lysate-resin is then transferred to a gravity column. Purifications took place in a 4°C cold room. Purification began by collecting lysate flowthrough, which was added back for a second pass through the column. The resin was then washed with 50 mL of the same buffer but with 30 mM imidazole. The proteins were eluted with 10 mL of buffer containing 250 mM imidazole. Eluted proteins were then dialyzed into storage buffer containing 50% glycerol, 50 mM Tris-HCl pH 7.5, 50 mM sodium chloride, 5 mM β -mercaptoethanol. Under these conditions, typical purifications would yield ~10 mg of protein per liter.

E. coli GlnRS and Hybrid Purification

E. coli GlnRS and hybrids genes were encoded in the expression plasmid, pSJW1. Each construct contained an N-terminal gIII periplasmic targeting tag and C-terminal His7 tag. The periplasmic does not get cleaved during expression but neither N- or C-terminal tag influence enzymatic activities. Plasmids were stored and propagated in Novablue cloning cell lines. Protein preps began each time by carrying out fresh transformations into BL21 DE3 pLysS expression cells under ampicillin and chloramphenicol selection. One liter cultures of LB media were inoculated with

saturated overnights and grown till reaching OD600 ~0.4. Expression began by the addition of 1.0 mM IPTG and grown for four hours at 37°C. Cells were then harvested by centrifugation at 4000 rpm for 15 minutes.

Lysates were prepared by resuspending the cell pellet in 20 mL of buffer containing 20 mM HEPES pH 7.5, 500 mM sodium chloride, 2 mM β -mercaptoethanol, 5 mM imidazole, 20 mg lysozyme, one Roche protease inhibitor tablet (no EDTA), 20 μ L of 10 mg/mL DNase1 and 10 μ L of 10 mg/mL RNase A. After pellets were resuspended they were lysed by sonication. Clarified lysates were prepared by centrifugation for 45 minutes at 15,000 rpm. Clarified lysates were then filtered through a 0.45 μ m syringe filter and then allowed to incubate with Ni-NTA resin for one hour at 4°C in a 50 mL falcon tube. The lysate-resin is then transferred to a gravity column. Purifications took place in a 4°C cold room. Purification began by collecting lysate flow through, which was added back for a second pass through the column. The resin was then washed to remove non-specific binding with two twenty-column volumes of buffer with increasing imidazole concentrations, 10 mM and 30 mM, respectively. Protein was then eluted with buffer containing 250 mM imidazole. Proteins were dialyzed and stored in 50 mM potassium phosphate pH 7.4, 50 mM sodium chloride, 5 mM β -mercaptoethanol and 50% glycerol. At this scale, purifications would yield ~2 mg of protein.

Enzyme kinetics experiments

Hand pipetting steady-state reactions

All steady state reactions were carried out by hand-pipetting and systematically were set at a final volume of 20 μL . Reactions were setup by mixing buffers, reducing agent, magnesium chloride and water separately. A 50-50 mix of unlabeled 10x tRNA and labeled tRNA were set mixed separately as well. The buffer and tRNA were individually heated to the desired temperature and mixed together such that the tRNA could refold with buffer containing magnesium ions. Folding conditions would need to be optimized for each system but typical variables include, denaturing time and temperature and rate of refolding time, specific details for the reported experiments are below. After tRNA and buffer equilibrate, amino acid and ATP are added resulting in 18 μL total volume. If reactions are carried out at 37°C, they would be pre-warmed for two minutes in a water bath before being initiated with 2 μL of 10x aaRS stock (tRNA is now at 1x). Reactions could be maintained in steady state by ensuring unlabeled tRNA is kept at least ten-fold higher than that of the final aaRS concentration. It is important to maintain the same volume of labeled tRNA for each reaction for equal signal to noise ratio, this includes when measuring $K_m[\text{tRNA}]$.

Time points are taken by removing 2 μL from the reaction and quenching into a solution containing 8 μL of 400 mM sodium acetate (pH 5.2) and 0.5 units of nuclease P1. Quenched time points are then digested at room temperature for at least 15 minutes. After digestion, 1.5 μL of the quenched/digested time point sample is spotted onto a PEI cellulose TLC plate. The TLC is then developed into the appropriate developing solution. For reactions using the amino acids glutamate and glutamine, this solution is 100 mM ammonium acetate and 5% acetic acid.

$K_M[\text{tRNA}]$, $K_M[\text{GLX}]$ and k_{cat} can be extracted from these experiments. After quantitating the fraction of product, the amount of $\text{GLX-tRNA}^{\text{Glx}}$ formed can be calculated based on the initial tRNA concentration. Measuring the slope of aminoacylation product formed computes the initial velocity. Initial velocities are then replotted on a Michaelis-Menton to obtain $V_{\max}$ and K_M .

- *Reaction buffers for E. coli GlnRS and hybrid variants:* 20 mM Tris-HCl pH 7.5, 10 mM magnesium chloride, 5 mM dithiothritol, 10 mM ATP. tRNAs were refolded by heating at 80°C for three minutes, mixed with reaction buffer and immediately set at ambient temperature but equilibrated for at least 15 minutes.

- *Reaction buffers for S. cerevisiae GlnRS and GluRS:* 50 mM Tris-HCl pH 7.5, 10 mM magnesium chloride, 10 mM β -mercaptoethanol, and 5 mM ATP. tRNAs were refolded by heating at 85°C for three minutes, mixed with reaction buffer and slow cooled in the heat block till reaching ambient temperature.

Rapid quench pre-steady state reactions

For pre-steady state kinetic experiments, these reactions will typically be carried out in the KinTek Rapid Quench. The final concentrations for buffers, ATP and amino acid will be identical to hand-pipetting experiments, but setup differently. The rapid quench has two sides; one can be kept radioactive. A typical time course requires eight time points and therefore requires 150 μL per time course (each time point in the reaction takes 15 μL plus dead volume within in the instrumentation). For pre-steady state

aminoacylation experiments use two 1 mL syringes: one containing the molar excess enzyme at two-times the final reaction concentration and the other contains the labeled tRNA, which is kept trace relative to enzyme concentration. Each syringe, as stated will hold 150 μ L with the appropriate and equal concentrations of buffers, magnesium chloride, reducing agents, ATP and amino acid. The buffers, ATP and amino acid concentration are kept equal and enzyme at 2x because each time point rapidly mixes a 50:50 ratio of the enzyme and tRNA syringe together. The rapid quench has a separate syringe holding 10 mL of quench solution. For pre-steady state aminoacylation experiments, the quench solution is 400 mM sodium acetate (pH 5.2) and 0.1% SDS (note at this time, not nuclease is added). SDS is added to the quench to ensure full quenching and helps resolve origins on the TLC due to higher protein concentration in these reactions.

As time points are taken, each results in a final quenched reaction sample results in $\sim 100 \mu$ L, 9 μ L of each time point is aliquoted into a 96-well plate. From here 0.5 units of nuclease P1 is added into each well and digested at room temperature for 15 minutes. From here, TLC work up is identical as described in steady state experiments.

$K_d[\text{tRNA}]$, $K_d[\text{GLX}]$ and $k_{obs,max}$ can be extracted from these experiments. This is accomplished by first quantitating the fraction of product formation over time using the phosphorimaging scans. Data can then be fit to first order exponential equation $[y=y_0(1-e^{-k_{obs} \cdot x})]$. Replotting k_{obs} against varied substrate concentration to a Michaelis-Menton curve results in a $K_d[\text{substrate}]$ and k_{max} .

Pulse-chase tRNA k_{off} experiments using KinTek Rapid Quench

The ^{32}P labeled tRNA based assay can also be applied to a Pulse-Chase styled experiment to measure $k_{\text{off}}[\text{tRNA}]$ for the aaRS, I did this for *S. cerevisiae* GlnRS and ΔNTD variant. The basic concept for this experiment is to incubate a saturating concentration of GlnRS and glutamine with trace amount of labeled tRNA in standard reaction buffer but *without* ATP. The goal is to maximize the amount of labeled *tRNA into a complex at equilibrium. Then allow 50-fold molar excess of unlabeled tRNA to incubate with the labeled GlnRS-*tRNA complex. The incubation time will vary from T_0 to T_{∞} . The concept is, as incubation time increases, any unbound enzyme will be sequestered and bind to the unlabeled tRNA (chase). To monitor the remaining bound GlnRS-labeled tRNA, the incubation time is “quenched” by at the addition of ATP to allow the aminoacylation reaction to move forward. The aminoacylation reaction goes to completion and then quenched with addition of 400 mM sodium acetate (pH 5.2) and 0.1% SDS. To obtain the k_{off} value, plot the fraction of product formation against incubation time and fit to a first order exponential decay.

Bioinformatics

Protein and tRNA sequences were obtained from the NCBI databases. tRNA sequences were also obtained from tRNAdb 2009 and Genomic tRNA Databases. Sequence database files were collated and aligned using eBioX. Structure based sequence alignments were performed using the MultiSeq package within VMD (Visual

Molecular Dynamics) software suite. To assist in mutation designs and generate figures, Schrödinger Pymol was used to visualize protein-RNA interactions.

References

1. Perona, J.J. and A. Hadd, *Structural diversity and protein engineering of the aminoacyl-tRNA synthetases*. Biochemistry, 2012. **51**(44): p. 8705-29.
2. Phizicky, E.M. and A.K. Hopper, *tRNA biology charges to the front*. Genes Dev, 2010. **24**(17): p. 1832-60.
3. Giege, R., M. Sissler, and C. Florentz, *Universal rules and idiosyncratic features in tRNA identity*. Nucleic Acids Res, 1998. **26**(22): p. 5017-35.
4. Madore, E., et al., *Effect of modified nucleotides on Escherichia coli tRNA^{Glu} structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34*. Eur J Biochem, 1999. **266**(3): p. 1128-35.
5. Rodriguez-Hernandez, A., et al., *Structural and mechanistic basis for enhanced translational efficiency by 2-thiouridine at the tRNA anticodon wobble position*. J Mol Biol, 2013. **425**(20): p. 3888-906.
6. Sylvers, L.A., et al., *A 2-thiouridine derivative in tRNA^{Glu} is a positive determinant for aminoacylation by Escherichia coli glutamyl-tRNA synthetase*. Biochemistry, 1993. **32**(15): p. 3836-41.
7. Yokoyama, S., et al., *Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon*. Proc Natl Acad Sci U S A, 1985. **82**(15): p. 4905-9.
8. O'Donoghue, P. and Z. Luthey-Schulten, *On the evolution of structure in aminoacyl-tRNA synthetases*. Microbiol Mol Biol Rev, 2003. **67**(4): p. 550-73.
9. O'Donoghue, P., et al., *The evolutionary history of Cys-tRNA^{Cys} formation*. Proc Natl Acad Sci U S A, 2005. **102**(52): p. 19003-8.
10. Perona, J.J. and Y.M. Hou, *Indirect readout of tRNA for aminoacylation*. Biochemistry, 2007. **46**(37): p. 10419-32.
11. Hauenstein, S., et al., *Shape-selective RNA recognition by cysteinyl-tRNA synthetase*. Nat Struct Mol Biol, 2004. **11**(11): p. 1134-41.
12. Sekine, S., et al., *Major identity determinants in the "augmented D helix" of tRNA^(Glu) from Escherichia coli*. J Mol Biol, 1996. **256**(4): p. 685-700.
13. de Duve, C., *Transfer RNAs: the second genetic code*. Nature, 1988. **333**(6169): p. 117-8.
14. Wiltrout, E., et al., *Misacylation of tRNA with methionine in Saccharomyces cerevisiae*. Nucleic Acids Res, 2012. **40**(20): p. 10494-506.
15. Liu, C., et al., *Kinetic quality control of anticodon recognition by a eukaryotic aminoacyl-tRNA synthetase*. J Mol Biol, 2007. **367**(4): p. 1063-78.
16. Guo, M., X.L. Yang, and P. Schimmel, *New functions of aminoacyl-tRNA synthetases beyond translation*. Nat Rev Mol Cell Biol, 2010. **11**(9): p. 668-74.
17. Havrylenko, S., et al., *Caenorhabditis elegans evolves a new architecture for the multi-aminoacyl-tRNA synthetase complex*. J Biol Chem, 2011. **286**(32): p. 28476-87.
18. Karanasios, E. and G. Simos, *Building arks for tRNA: structure and function of the Arc1p family of non-catalytic tRNA-binding proteins*. FEBS Lett, 2010. **584**(18): p. 3842-9.
19. Lee, S.W., et al., *Aminoacyl-tRNA synthetase complexes: beyond translation*. J Cell Sci, 2004. **117**(Pt 17): p. 3725-34.
20. Cestari, I., et al., *A multiple aminoacyl-tRNA synthetase complex that enhances tRNA-aminoacylation in African trypanosomes*. Mol Cell Biol, 2013. **33**(24): p. 4872-88.
21. Ko, Y.G., et al., *Glutamine-dependent antiapoptotic interaction of human glutaminyl-tRNA synthetase with apoptosis signal-regulating kinase 1*. J Biol Chem, 2001. **276**(8): p. 6030-6.

22. Bertram, P.G., et al., *Convergence of TOR-nitrogen and Snf1-glucose signaling pathways onto Gln3*. Mol Cell Biol, 2002. **22**(4): p. 1246-52.
23. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation*. Science, 2009. **324**(5930): p. 1029-33.
24. Zhang, Y. and J.M. Yang, *Altered energy metabolism in cancer: a unique opportunity for therapeutic intervention*. Cancer Biol Ther, 2013. **14**(2): p. 81-9.
25. Davis, L. and J.W. Chin, *Designer proteins: applications of genetic code expansion in cell biology*. Nat Rev Mol Cell Biol, 2012. **13**(3): p. 168-82.
26. Neumann, H., *Rewiring translation - Genetic code expansion and its applications*. FEBS Lett, 2012. **586**(15): p. 2057-64.
27. Siatecka, M., et al., *Modular evolution of the Glx-tRNA synthetase family--rooting of the evolutionary tree between the bacteria and archaea/eukarya branches*. Eur J Biochem, 1998. **256**(1): p. 80-7.
28. Sheppard, K. and D. Soll, *On the evolution of the tRNA-dependent amidotransferases, GatCAB and GatDE*. J Mol Biol, 2008. **377**(3): p. 831-44.
29. Hadd, A.a.P., J.J., *Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl-tRNA synthetases*. Journal of Molecular Biology, 2014. **in the press**.
30. Chang, K.M. and T.L. Hendrickson, *Recognition of tRNAGln by Helicobacter pylori GluRS2--a tRNAGln-specific glutamyl-tRNA synthetase*. Nucleic Acids Res, 2009. **37**(20): p. 6942-9.
31. Whelihan, E.F. and P. Schimmel, *Rescuing an essential enzyme-RNA complex with a non-essential appended domain*. EMBO J, 1997. **16**(10): p. 2968-74.
32. Grant, T.D., et al., *Structural conservation of an ancient tRNA sensor in eukaryotic glutaminyl-tRNA synthetase*. Nucleic Acids Res, 2012. **40**(8): p. 3723-31.
33. Oshikane, H., et al., *Structural basis of RNA-dependent recruitment of glutamine to the genetic code*. Science, 2006. **312**(5782): p. 1950-4.
34. Perona, J.J., M.A. Rould, and T.A. Steitz, *Structural basis for transfer RNA aminoacylation by Escherichia coli glutaminyl-tRNA synthetase*. Biochemistry, 1993. **32**(34): p. 8758-71.
35. Rodriguez-Hernandez, A. and J.J. Perona, *Heat maps for intramolecular communication in an RNP enzyme encoding glutamine*. Structure, 2011. **19**(3): p. 386-96.
36. Sherlin, L.D. and J.J. Perona, *tRNA-dependent active site assembly in a class I aminoacyl-tRNA synthetase*. Structure, 2003. **11**(5): p. 591-603.
37. Uter, N.T., I. Gruic-Sovulj, and J.J. Perona, *Amino acid-dependent transfer RNA affinity in a class I aminoacyl-tRNA synthetase*. J Biol Chem, 2005. **280**(25): p. 23966-77.
38. Ibba, M., et al., *Interactions between tRNA identity nucleotides and their recognition sites in glutaminyl-tRNA synthetase determine the cognate amino acid affinity of the enzyme*. Proc Natl Acad Sci U S A, 1996. **93**(14): p. 6953-8.
39. Uter, N.T. and J.J. Perona, *Long-range intramolecular signaling in a tRNA synthetase complex revealed by pre-steady-state kinetics*. Proc Natl Acad Sci U S A, 2004. **101**(40): p. 14396-401.
40. Perona, J.J., et al., *Structural basis for misaminoacylation by mutant E. coli glutaminyl-tRNA synthetase enzymes*. Science, 1989. **246**(4934): p. 1152-4.
41. Rath, V.L., et al., *How glutaminyl-tRNA synthetase selects glutamine*. Structure, 1998. **6**(4): p. 439-49.
42. Uter, N.T. and J.J. Perona, *Active-site assembly in glutaminyl-tRNA synthetase by tRNA-mediated induced fit*. Biochemistry, 2006. **45**(22): p. 6858-65.
43. Corigliano, E.M. and J.J. Perona, *Architectural underpinnings of the genetic code for glutamine*. Biochemistry, 2009. **48**(4): p. 676-87.

44. Bullock, T.L., et al., *Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants*. J Mol Biol, 2003. **328**(2): p. 395-408.
45. Grant, T.D., et al., *Structural conservation of an ancient tRNA sensor in eukaryotic glutamyl-tRNA synthetase*. Nucleic Acids Res, 2011.
46. Rodriguez-Hernandez, A., et al., *Synthesis of Glu-tRNA(Gln) by engineered and natural aminoacyl-tRNA synthetases*. Biochemistry, 2010. **49**(31): p. 6727-36.
47. Ibba, M. and D. Soll, *Aminoacyl-tRNAs: setting the limits of the genetic code*. Genes Dev, 2004. **18**(7): p. 731-8.
48. Mirande, M., *Processivity of translation in the eukaryote cell: role of aminoacyl-tRNA synthetases*. FEBS Lett, 2010. **584**(2): p. 443-7.
49. Guo, M. and P. Schimmel, *Essential nontranslational functions of tRNA synthetases*. Nat Chem Biol, 2013. **9**(3): p. 145-53.
50. Jia, J., et al., *WHEP domains direct noncanonical function of glutamyl-Prolyl tRNA synthetase in translational control of gene expression*. Mol Cell, 2008. **29**(6): p. 679-90.
51. Xu, X., et al., *Unique domain appended to vertebrate tRNA synthetase is essential for vascular development*. Nat Commun, 2012. **3**: p. 681.
52. Kyriacou, S.V. and M.P. Deutscher, *An important role for the multienzyme aminoacyl-tRNA synthetase complex in mammalian translation and cell growth*. Mol Cell, 2008. **29**(4): p. 419-27.
53. Negrutskii, B.S. and M.P. Deutscher, *Channeling of aminoacyl-tRNA for protein synthesis in vivo*. Proc Natl Acad Sci U S A, 1991. **88**(11): p. 4991-5.
54. Salazar, J.C., et al., *Coevolution of an aminoacyl-tRNA synthetase with its tRNA substrates*. Proc Natl Acad Sci U S A, 2003. **100**(24): p. 13863-8.
55. Tumbula, D.L., et al., *Domain-specific recruitment of amide amino acids for protein synthesis*. Nature, 2000. **407**(6800): p. 106-10.
56. Lapointe, J., L. Duplain, and M. Proulx, *A single glutamyl-tRNA synthetase aminoacylates tRNA^{Glu} and tRNA^{Gln} in Bacillus subtilis and efficiently misacylates Escherichia coli tRNA^{Gln1} in vitro*. J Bacteriol, 1986. **165**(1): p. 88-93.
57. Wilcox, M. and M. Nirenberg, *Transfer RNA as a cofactor coupling amino acid synthesis with that of protein*. Proc Natl Acad Sci U S A, 1968. **61**(1): p. 229-36.
58. Curnow, A.W., et al., *Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation*. Proc Natl Acad Sci U S A, 1997. **94**(22): p. 11819-26.
59. Woese, C.R., et al., *Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process*. Microbiol Mol Biol Rev, 2000. **64**(1): p. 202-36.
60. Bullock, T.L., et al., *A rationally engineered misacylating aminoacyl-tRNA synthetase*. Proc Natl Acad Sci U S A, 2008. **105**(21): p. 7428-33.
61. O'Donoghue, P., et al., *Rational design of an evolutionary precursor of glutamyl-tRNA synthetase*. Proc Natl Acad Sci U S A, 2011. **108**(51): p. 20485-90.
62. Hayase, Y., et al., *Recognition of bases in Escherichia coli tRNA(Gln) by glutamyl-tRNA synthetase: a complete identity set*. EMBO J, 1992. **11**(11): p. 4159-65.
63. Frugier, M., et al., *Identity switches between tRNAs aminoacylated by class I glutamyl- and class II aspartyl-tRNA synthetases*. Biochemistry, 1994. **33**(33): p. 9912-21.
64. Rogers, K.C. and D. Soll, *Discrimination among tRNAs intermediate in glutamate and glutamine acceptor identity*. Biochemistry, 1993. **32**(51): p. 14210-9.
65. Rould, M.A., et al., *Structure of E. coli glutamyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution*. Science, 1989. **246**(4934): p. 1135-42.
66. Rould, M.A., J.J. Perona, and T.A. Steitz, *Structural basis of anticodon loop recognition by glutamyl-tRNA synthetase*. Nature, 1991. **352**(6332): p. 213-8.
67. Sekine, S., et al., *ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding*. EMBO J, 2003. **22**(3): p. 676-88.

68. Nureki, O., et al., *Structure of an archaeal non-discriminating glutamyl-tRNA synthetase: a missing link in the evolution of Gln-tRNA^{Gln} formation*. Nucleic Acids Res, 2010. **38**(20): p. 7286-97.
69. Sekine, S., et al., *Structural basis for anticodon recognition by discriminating glutamyl-tRNA synthetase*. Nat Struct Biol, 2001. **8**(3): p. 203-6.
70. Kim, T., et al., *Catalytic peptide of human glutamyl-tRNA synthetase is essential for its assembly to the aminoacyl-tRNA synthetase complex*. J Biol Chem, 2000. **275**(28): p. 21768-72.
71. Wang, C.C. and P. Schimmel, *Species barrier to RNA recognition overcome with nonspecific RNA binding domains*. J Biol Chem, 1999. **274**(23): p. 16508-12.
72. Deniziak, M., et al., *Deinococcus glutamyl-tRNA synthetase is a chimera between proteins from an ancient and the modern pathways of aminoacyl-tRNA formation*. Nucleic Acids Res, 2007. **35**(5): p. 1421-31.
73. Simos, G., et al., *The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl- and glutamyl-tRNA synthetases*. EMBO J, 1996. **15**(19): p. 5437-48.
74. Simader, H., et al., *Structural basis of yeast aminoacyl-tRNA synthetase complex formation revealed by crystal structures of two binary sub-complexes*. Nucleic Acids Res, 2006. **34**(14): p. 3968-79.
75. Graindorge, J.S., et al., *Role of Arc1p in the modulation of yeast glutamyl-tRNA synthetase activity*. Biochemistry, 2005. **44**(4): p. 1344-52.
76. Frechin, M., et al., *Yeast mitochondrial Gln-tRNA(Gln) is generated by a GatFAB-mediated transamidation pathway involving Arc1p-controlled subcellular sorting of cytosolic GluRS*. Genes Dev, 2009. **23**(9): p. 1119-30.
77. Grant, T.D., et al., *The structure of yeast glutamyl-tRNA synthetase and modeling of its interaction with tRNA*. J Mol Biol, 2013. **425**(14): p. 2480-93.
78. Johnson, K.A., *Advances in transient-state kinetics*. Curr Opin Biotechnol, 1998. **9**(1): p. 87-9.
79. Karanasios, E., et al., *Molecular determinants of the yeast Arc1p-aminoacyl-tRNA synthetase complex assembly*. J Mol Biol, 2007. **374**(4): p. 1077-90.
80. Hans, M.A., E. Heinzle, and C. Wittmann, *Quantification of intracellular amino acids in batch cultures of Saccharomyces cerevisiae*. Appl Microbiol Biotechnol, 2001. **56**(5-6): p. 776-9.
81. Sissler, M., et al. *Mitochondrial Aminoacyl-tRNA Synthetases*. Madame Curie Bioscience Database 2000.
82. Deinert, K., et al., *Arc1p organizes the yeast aminoacyl-tRNA synthetase complex and stabilizes its interaction with the cognate tRNAs*. J Biol Chem, 2001. **276**(8): p. 6000-8.
83. Simader, H., M. Hothorn, and D. Suck, *Structures of the interacting domains from yeast glutamyl-tRNA synthetase and tRNA-aminoacylation and nuclear-export cofactor Arc1p reveal a novel function for an old fold*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 12): p. 1510-9.
84. Bridgham, J.T., E.A. Ortlund, and J.W. Thornton, *An epistatic ratchet constrains the direction of glucocorticoid receptor evolution*. Nature, 2009. **461**(7263): p. 515-9.
85. Juhling, F., et al., *tRNAdb 2009: compilation of tRNA sequences and tRNA genes*. Nucleic Acids Res, 2009. **37**(Database issue): p. D159-62.
86. Cheng, A.A. and T.K. Lu, *Synthetic biology: an emerging engineering discipline*. Annu Rev Biomed Eng, 2012. **14**: p. 155-78.
87. Chin, J.W., *Expanding and reprogramming the genetic code of cells and animals*. Annu Rev Biochem, 2014. **83**: p. 379-408.
88. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code*. Annu Rev Biochem, 2010. **79**: p. 413-44.

89. O'Donoghue, P., et al., *Upgrading protein synthesis for synthetic biology*. Nat Chem Biol, 2013. **9**(10): p. 594-8.
90. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome*. Nature, 2010. **464**(7287): p. 441-4.
91. Wang, L., et al., *Expanding the genetic code of Escherichia coli*. Science, 2001. **292**(5516): p. 498-500.
92. Chin, J.W., et al., *An expanded eukaryotic genetic code*. Science, 2003. **301**(5635): p. 964-7.
93. Wang, L. and P.G. Schultz, *A general approach for the generation of orthogonal tRNAs*. Chem Biol, 2001. **8**(9): p. 883-90.
94. Liu, D.R., et al., *Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo*. Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10092-7.
95. Fan, C., et al., *Exploring the Substrate Range of Wild-Type Aminoacyl-tRNA Synthetases*. Chembiochem, 2014.
96. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding N(epsilon)-acetyllysine in recombinant proteins*. Nat Chem Biol, 2008. **4**(4): p. 232-4.
97. Mukai, T., et al., *Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases*. Biochem Biophys Res Commun, 2008. **371**(4): p. 818-22.
98. Nehring, S., N. Budisa, and B. Wiltschi, *Performance analysis of orthogonal pairs designed for an expanded eukaryotic genetic code*. PLoS One, 2012. **7**(4): p. e31992.
99. Guo, L.T., et al., *Rational design and directed evolution of a bacterial-type glutaminyl-tRNA synthetase precursor*. Nucleic Acids Res, 2012.
100. Dulic, M., et al., *Partitioning of tRNA-dependent editing between pre- and post-transfer pathways in class I aminoacyl-tRNA synthetases*. J Biol Chem, 2010. **285**(31): p. 23799-809.
101. Hedstrom, L., L. Szilagyi, and W.J. Rutter, *Converting trypsin to chymotrypsin: the role of surface loops*. Science, 1992. **255**(5049): p. 1249-53.
102. Bennett, B.D., et al., *Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli*. Nat Chem Biol, 2009. **5**(8): p. 593-9.
103. Francklyn, C.S., et al., *Methods for kinetic and thermodynamic analysis of aminoacyl-tRNA synthetases*. Methods, 2008. **44**(2): p. 100-18.
104. Simos, G., et al., *A conserved domain within Arc1p delivers tRNA to aminoacyl-tRNA synthetases*. Mol Cell, 1998. **1**(2): p. 235-42.
105. Frechin, M., et al., *Arc1p: anchoring, routing, coordinating*. FEBS Lett, 2010. **584**(2): p. 427-33.
106. Kwon, H.S., et al., *Identification of CD23 as a functional receptor for the proinflammatory cytokine AIMP1/p43*. J Cell Sci, 2012. **125**(Pt 19): p. 4620-9.
107. Park, S.G., et al., *Hormonal activity of AIMP1/p43 for glucose homeostasis*. Proc Natl Acad Sci U S A, 2006. **103**(40): p. 14913-8.
108. Rolland, F., J. Winderickx, and J.M. Thevelein, *Glucose-sensing mechanisms in eukaryotic cells*. Trends Biochem Sci, 2001. **26**(5): p. 310-7.
109. Sherlin, L.D., et al., *Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization*. RNA, 2001. **7**(11): p. 1671-8.
110. Fechter, P., et al., *Ribozyme processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: production and activity*. FEBS Lett, 1998. **436**(1): p. 99-103.
111. Wolfson, A.D. and O.C. Uhlenbeck, *Modulation of tRNAAla identity by inorganic pyrophosphatase*. Proc Natl Acad Sci U S A, 2002. **99**(9): p. 5965-70.

112. Suzuki, T. and T. Suzuki, *Chaplet column chromatography: isolation of a large set of individual RNAs in a single step*. Methods Enzymol, 2007. **425**: p. 231-9.