URZYMES – MINIMAL AMINOACYL-TRNA SYNTHETASES: ADDRESSING THEIR CATALYTIC AUTHENTICITY AND TECHNICAL CHALLENGES TO UTILIZATION

Jessica Jean Hobson

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

Chapel Hill
2022

Approved by:

Charles W. Carter, Jr.

Qi Zhang

Christina Burch

Saskia Neher

Ronald Swanstrom

ABSTRACT

Jessica Jean Hobson: Urzymes – Minimal Aminoacyl-tRNA Synthetases: Addressing Their Catalytic Authenticity and Technical Challenges to Utilization
(Under the direction of Charles W. Carter, Jr.)

Aminoacyl-tRNA synthetase (aaRS)/tRNA cognate pairs translate the genetic code by synthesizing specific aminoacyl-tRNAs that are assembled on messenger RNA by the ribosome. Deconstruction of various members of the two distinct aaRS superfamilies (Classes) has provided conceptual and experimental models for their early evolution. Urzymes, containing ~120-130 amino acids excerpted from regions where sequence complementarities have been identified, are key experimental models motivated by the proposal of a single bidirectional ancestral gene. Previous reports have claimed that urzymes from Class I and Class II accelerate both amino acid activation and tRNA aminoacylation. However, some of that work has proven difficult to reproduce. A chimeric tryptophanyl-tRNA synthetase urzyme had enhanced solubility but lacked catalytic activity. To further investigate the authenticity of catalysis by urzymes, we have examined a third urzyme (LeuAC) prepared from the Class I *Pyrococcus horikoshii* Leucyl-tRNA synthetase. We (i) summarize difficulties encountered in authenticating its two catalytic functions, amino acid activation and tRNA aminoacylation; (ii) adduce evidence for the authenticity of both canonical reactions; and (iii) describe a prominent non-canonical enzymatic function—production of the product ADP—that has been overlooked since it first was described for several full-length aaRS in the 1970s.

To my dearest Jaxon, you have brought me so much love and joy throughout the years.
You are my motivation, my happiness, and my greatest achievement.
I will love you forever – Mom

ACKNOWLEDGEMENTS

I would like to thank my mentor Charlie for being understanding, accommodating, and supportive during all my years in the lab. I have had an unusual graduate career, with many potentially derailing obstacles, but with his help and support I succeeded. My committee members also showed great care and patience during my time at UNC and I would like to thank them, especially Sasky, who provided a lot of external support for both me and Jaxon. I would also like to thank Derek Mann, Adam Burch, and Holly Simmons for their time, effort, and dedication. I am grateful to have had assistance from each of you and am so very proud of what you each accomplished. I would like to thank Tishan Williams who helped to get me started in the lab and continues to provide invaluable guidance and friendship Aaron Cook provided many laughs and an enjoyable work environment that was greatly missed upon his leaving the lab – thank you. I must also thank Zhijie Li for all his help, support, and commiseration throughout our years together. We shared a lot of struggles and faced many obstacles, and I am glad to have faced them with you – thank you again. I would like to thank Lynn Ray for her unwavering support and many much-needed chats during my time at UNC. In addition, would like to thank all my friends and peers who have helped support me both in and out of the lab, especially Shawna Daniels and Cassandra Hayne; both contributed to my success in different ways, and I am grateful for their help. Finally, I must thank my family. All of you filled my head with boundless hopes and dreams and always believed in me, even when I didn't. You each have helped to keep me on track in your own special way and provided endless love and support. I love each and every one of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

β-ME	Beta-Mercaptoethanol
aaRS	Aminoacyl-tRNA synthetase
ADP	Adenosine diphosphate
AlaRS	Alinyl-tRNA synthetase
AMP	Adenosine monophosphate
ArgRS	Arginyl-tRNA synthetase
AsnRS	Asparaginyl-tRNA synthetase
AspRS	Aspartyl-tRNA synthetase
AST	Active-site titration
ATP	Adenosine triphosphate
CP1	Connecting peptide 1
CP1	Connecting peptide 2
CysRS	Cystidyl-tRNA synthetase
D146A	Aspartic acid 146 to alanine
DTT	Dithiothreitol
EB	Evans blue
EDTA	Ethylenediaminetetraacetic acid
EKK	Glutamic acid-Lysine-Lysine
GlnRS	Glutanyl-tRNA synthetase
GluRS	Glutamyl-tRNA synthetase
GlyRS	Glycyl-tRNA synthetase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)

HisRS	Histidyl-tRNA synthetase
IleRS	Isoleucyl-tRNA synthetase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KCl	Potassium chloride
KF	Potassium flouride
LeuAC	Leucyl-tRNA synthetase urzyme
LeuRS	Leucyl-tRNA synthetase
LysRS	Lysyl-tRNA synthetase
$MgCl_2$	Magnesium chloride
MBP	Maltose-binding protein
MetRS	Methionyl-tRNA synthetase
N63D	Asparagine 63 to aspartic acid
N63E	Asparagine 63 to glutamic acid
NaCl	Sodium chloride
Ni-NTA	Nickel nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OB25GH	Optimal buffer 25 guanidine hydrochloride
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PheRS	Phenylalanyl-tRNA synthetase
PNK	Polynucleotide kinase
PPi	Pyrophosphate
ProRS	Prolinyl-tRNA synthetase

RO	Rodin Ohno
SDS	Sodium dodecyl sulfate
SerRS	Serinyl-tRNA synthetase
TCA	Trichloroacetic acid
ThrRS	Threonyl-tRNA synthetase
TEV	Tobacco Etch virus
Tris	Tris(hydroxymethyl)aminomethane
TrpAC	Tryptophanyl-tRNA synthetase urzyme
TrpRS	Tryptophanyl-tRNA synthetase
TyrRS	Tyrosyl-tRNA synthetase
ValRS	Valinyl-tRNA synthetase

CHAPTER I: INTRODUCTION

The origin of life is the greatest remaining mystery in all of science. There are a handful of well-established theories, such as the RNA and RNA/peptide world hypotheses (1, 2), that seek to explain how small molecules self-organize to allow for reproduction with great enough fidelity to overcome error threshold (3). The key to understanding accurate self-replication lies in the origin of the genetic code, for this seemingly simple three-letter code likely allowed for evolution from a small overlapping pool of codons to the well-diverged codon code of extant life (2,4-8).

The two basic features needed to implement a genetic code include information storage and execution (2). In extant life, DNA or RNA store the genetic information. The execution of information is carried out by RNA and proteins. The two types of execution molecules can be further broken down into the specific roles they play. For example, RNA acts both as a structural component in the ribosome and an adapter molecule in the form of tRNA. While proteins too act as structural components, as in the ribosome, their prominent roles are as catalysts – like aminoacyl-tRNA synthetases (aaRSs) which activate amino acids (aa) and then attach them to [aminoacylate] tRNAs.

Aminoacyl-tRNA Synthetases

aaRSs act as the linchpin of genetic code by ensuring the fidelity of protein translation. These enzymes activate amino acids, making them chemically reactive, by the covalent addition of adenosine monophosphate, generating an unstable carboxylate-phosphate anhydride. The mixed anhydride is then attached to the 2’ or 3’ ribose hydroxyl of the tRNA 3’ adenosine –

tRNA aminoacylation. Furthermore, some aaRSs catalyze the hydrolysis of tRNAs misacylated with the incorrect amino acid (9).

There are 20 common aaRSs which are divided in two classes based on architecture of active-site. Class I (ArgRS, CysRS, GlnRS, GluRS, IleRS, LeuRS, MetRS, TrpRS, TyrRS, ValRS) synthetases contain a Rossmann fold, are generally monomeric, bind the minor groove of tRNA with the highly conserved PxxxxHIGH and KMSKS motifs, and acylate the ribose 2' hydroxyl (Figure 1A). Class II (AlaRS, AsnRS, AspRS, GlyRS, HisRS, LysRS, PheRS, ProRS, SerRS, ThrRS) synthetase active-sites contain antiparallel beta strands which use motifs I-III to bind the major groove of tRNA, acylating the ribose 3' hydroxyl, and are generally di- or multimeric (9) (Figure 1B). Amino acid classification follows the two-class division of aaRSs.

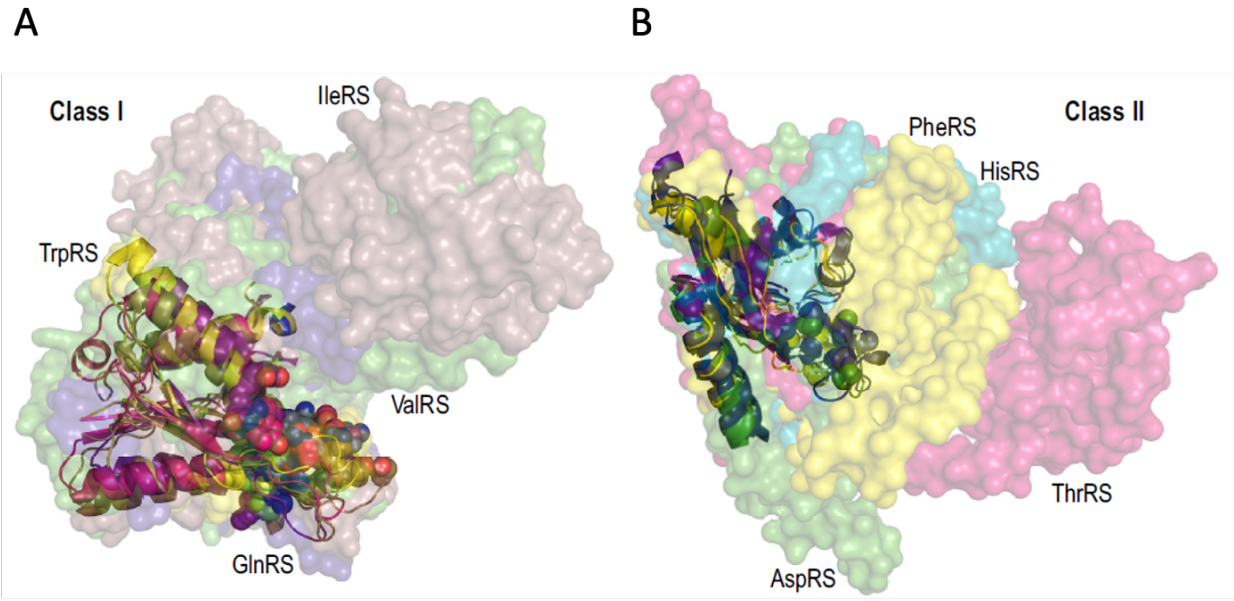


Figure 1. Class I and II aminoacyl-tRNA synthetase structures. A. Alignment of four Class I crystal structures (space-filling) reveals structurally conserved core (ribbon) composed of the Rossman fold and PxxxxHIGH and KMSKS motifs. B. Alignment of four Class II crystal structures (space-filling) reveals structurally conserved core (ribbon) of anti-parallel beta strands and motifs I-III. Originally published in (18) and reproduced with permission.

Rodin Ohno Hypothesis: Bi-directional coding

The absence of structural or sequence conservation between class I and II synthetases points to their independent evolution (10,11) which is incompatible with the singular origin of tRNAs (12) and the coevolution of each with the genetic code (13). Rodin and Ohno (14) observed that the highly conserved class I motifs PxxxxHIGH and KMSKS are complements of the conserved class II motifs I and II. They proposed that class I and II synthetases did coevolve – but as opposite strands of the same gene. This theory resolved the apparent inconsistencies of the coevolution of the two distinct and unrelated aaRS classes.

Urzymes: Resurrected Ancestral Aminoacyl-tRNA Synthetases

Experimental support for the Rodin Ohno hypothesis

Inspired by the Rodin and Ohno (RO) hypothesis, we set out to create and test the activity of a set of bi-directionally coded aaRSs (15,16). Briefly, 24 crystal structures of Class I synthetases were aligned and used to identify the non-conserved regions to be removed. The resulting conserved core was then used to derive a specific *Bacillus stearothermophilus* TrpRS core. The Rosetta software suite was used to close gaps from Connecting Peptide 1 (CP1) removal and then again to reduce the amount of exposed hydrophobic residues (Figure 2). The resulting designed TrpRS minimal catalytic domain (MCD) gene was purified and renatured from inclusion bodies and found to be catalytically active. This experiment provided proof-of-principle that a minimal aaRS, termed urzyme, could in fact represent a functional ancestral aaRS.

To complete testing of the RO hypothesis, the urzyme gene [derived from the wild-type (WT) *B. stearothermophilus* TrpRS] was aligned antiparallel to the *Escherichia coli* HisRS gene using GAP (17). The alignment placed the class I PxxxxHIGH motif antiparallel to the class II

motif II and the class I KMSKS motif antiparallel to class II motif I, with a 30 nt gap following the motif I of class II – this is the stretch of greatest sequence heterogeneity in class II (Figure 2). Of the 96 codons aligned, ~ 45% of them had complementary middle-bases which is well above the expected value of 25% for a random codon alignment. Four varying HisRS urzyme constructs were designed, purified, and tested for amino acid activation. All constructs demonstrated catalytic activity, thus supporting the RO hypothesis that aaRSs arose from opposite strands of a bi-directional gene.

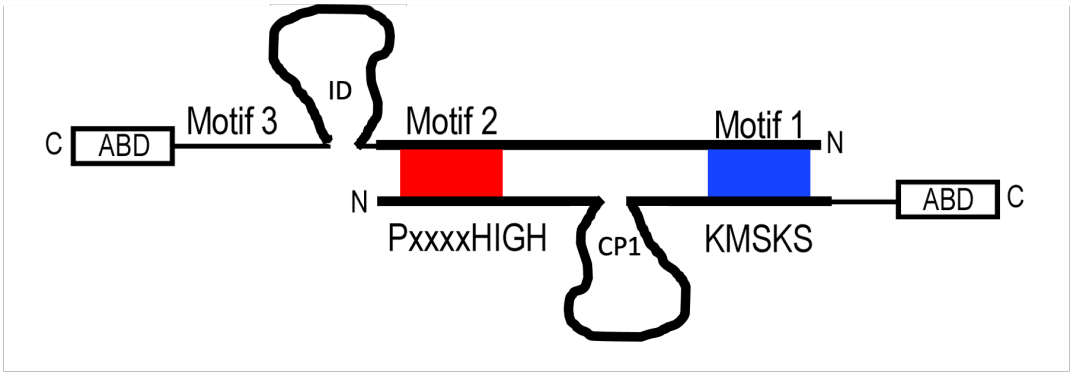


Figure 2. Schematic representation of Class I and II urzyme domain alignment. Class II synthetase domains are on top (C- to N-terminal: Anti-codon binding, Motif 3, Insertion, Motif 2, Motif 1) and Class I synthetase domains are on the bottom (C- to N-terminal: Anti-codon binding, KMSKS, Connecting Peptide 1, PxxxHIGH). Red and blue rectangles highlight the anti-parallel alignment of codons between Class I PxxxxHIGH and Class II Motif 2, and Class I KMSKS and Class II Motif 1, respectively.

Urzymes catalyze amino acid activation and aminoacylation

In addition to testing the ability of urzymes to catalyze amino acid activation, we also demonstrated their ability to aminoacylate tRNAs and determined their affinity for all three substrates (18). Interestingly, the urzymes exhibit different trends in their substrate affinities when compared to their full-length counterparts. For example, the TrpRS urzyme has a much-reduced affinity for tryptophan and tRNA as compared to the wild-type enzyme (tryptophan - 2.4 mM urzyme vs. 3.2 μM wild-type; tRNA – 7.49 μM urzyme vs. 0.57 μM wild-type) but has a

similar affinity for ATP (240 μM urzyme vs. 300 μM wild-type. In contrast, the HisRS urzyme has a higher affinity for ATP (11 μM urzyme vs. 890 μM wild-type), a lower affinity for histidine (260 μM vs. 30 μM wildtype), and a similar affinity for tRNA (4.9 μM both).

Urzyme nomenclature

From here on the discussion will pertain only to Class I urzymes and for convenience will use a nomenclature based on a four-module structure exhibited by all Class I aaRS. The N-terminal ATP-binding site containing the HIGH signature (the protozyme, ~46 residues) is denoted A, the variable-length insertion connecting peptide 1 (CP1) B, the second two thirds of the Rossmann dinucleotide fold (specificity helix and KMSKS signature) C, and the C-terminal anticodon-binding domain is denoted D (Figure 2) (15,19). With this nomenclature, the Class I urzymes will be called TrpAC and LeuAC. They contain only the highly conserved A and C modules that compose the catalytic machinery and lack the B and D fragments necessary for enhanced discrimination of amino acid and tRNA substrates.

Urzyme Derived Methodologies and Potential Applications

The methodology used to derive urzymes may prove useful in the development of novel proteins. Urzymology uses structural, instead of primary sequence, alignments to infer the root amino acid. Structural superposition can also tell us what amino acids may be used, instead of just what were used, and it gives us the ability to find ones that are better suited to new applications. Gene therapy could benefit from using similar methods to minimize therapeutic enzymes sequences and make delivery in viral vectors feasible for a broader range of genes. Moreover, urzymes themselves have potential application beyond origin of life research; owing

to their relaxed substrate specificity, urzymes have the protentional to be used for non-canonical amino acid incorporation into synthetic and therapeutic proteins.

Urzymes: Key to Unlocking the Origin of the Genetic Code and Coded Protein Translation

The way to understand the origin of the genetic code is to experimentally test hypotheses about its evolution. Urzymes offer a tool to test such hypotheses, as they represent aaRSs that were present during the origin of the genetic code. Consensus in the community is that there must have been a much-reduced genetic code in the beginning. The original translation machinery was likely much more promiscuous, then duplication events led to divergence and specialization of function. There are two distinct coding regions on extant tRNA – the anti-codon loop and acceptor stem. Together the two regions can discriminate size (acceptor) and polarity (anti-codon). The acceptor stem likely served as the first and only "anti-codon." The tRNA minihelix is believed to be the predecessor of tRNA which evolved due to a gene duplication event. Together urzymes and minihelix tRNAs could serve as an experimental model that could help unlock the secrets of the origin of the genetic code and coded protein translation.

CHAPTER II: GENERATING SOLUBLE AND PROFICIENT URZYMES

The original TrpRS urzyme was insoluble and had to be isolated from inclusion bodies, even when expressed as a MBP fusion protein. The refolding process proved very challenging to reproduce, so we tried two complementary methods to increase TrpAC solubility.

A novel LeuRS urzyme, first reported in (20) and further characterized here, was always more soluble than TrpAC and could be isolated from the soluble fraction when cells were lysed in an optimized buffer. Since both TrpAC and LeuAC are derived from Class I synthetases and share the same structural core, we decided to look for differences between the structures to identify what could be leading to the increased solubility of the LeuRS urzyme. We found LeuAC possessed an N-terminal extension of three charged residues and a salt-bridge at the rear-side of the Rossmann fold (Figure 4A). We hypothesized that these features could contribute to the increased solubility and/or stability of LeuAC and hence chose to add these features to the TrpRS urzyme. In addition, we also wanted to test the effect of the D146A mutation that was previously found to increase TrpAC activity.

We hypothesized that replacing TrpRS with a series of four successively less active TrpRS mutants would allow us to select for *E. coli* with background mutations that enable it to tolerate less functional TrpRSs (Figure 3). We predicted this progressive adaptation to decreased synthetase function would allow us to replace the final mutant TrpRS with TrpAC, as their free energies of amino acid activation are similar.

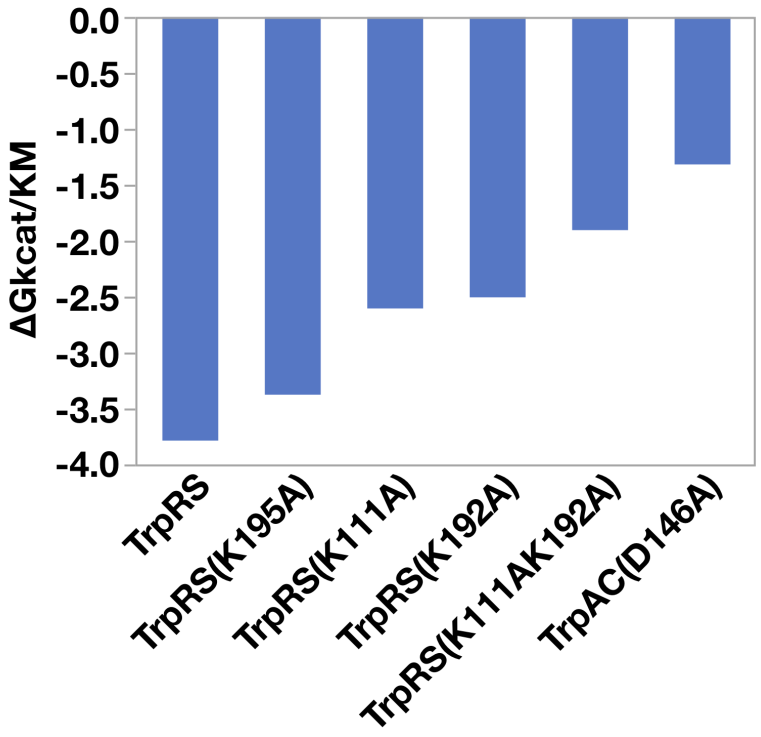


Figure 3. Strategy for progressive TrpRS evolution. The free energy of amino acid activation is shown on the Y-axis. From left to right the free energy increases with increasing mutational load from the wild-type enzyme to the TrpAC(D146A) urzyme.

Experimental Procedures

Grafting LeuAC features onto TrpAC

PCR was used to add three residues, EKK, after the start codon to a C-terminally hepta-histidine tagged TrpAC gene. The PCR product, EKK-TrpAC-6xHIS, was cloned into pBAD24 and transformed into the E. coli strain BW27783, to allow for linear protein induction from the pBAD (arabinose inducible) promoter (21). Mutagenic PCR using EKK-TrpAC-6xHIS as a template was performed to produce both the EKK-TrpAC(N63D)-6xHIS and EKK-TrpAC(N63E)-6xHIS alleles.

TrpAC solubility assay

For solubility experiments, overnight cultures were diluted in 500 ml LB plus 100 ng/ml ampicillin and grown at 37°C to an OD_{600} 0.4-0.6. Cultures were then induced with 0.2% arabinose and grown for an additional 2 at 37°C. Cells were harvested by centrifugation and stored at -80°C. Cells were resuspended in Optimal Buffer 25 guanidine hydrochloride (OB25GH) (20 mM Tris, pH 7.4, 1 mM EDTA, 5 mM β-ME, 17.5% glycerol, 0.5% NP40, 33 mM ammonium sulfate, 1.25% glycine, 300 mM guanidine hydrochloride) and lysed with an Avestin Emulsiflex c5. The soluble and insoluble fractions were separated by centrifugation. For solubility assays, the soluble fraction was precipitated with acetone to remove buffer. The pellet and precipitated soluble fraction were each resuspended in 1x protein loading dye to one-half the original cell lysis volume. Aliquots of the samples were separated on 4-20% SDS polyacrylamide gels (BioRad) and transferred to a nitrocellulose membrane. Western blot was performed with SuperSignal™ West HisProbe Kit following the manufacture's protocol. For activity assays, the proteins were purified from the soluble fraction using Ni-NTA affinity chromatography and eluted stepwise with 0.1 – 1 M imidazole. Fractions containing the protein of interest were pooled and concentrated. The fraction of active-sites that are competent for catalysis were determined using active-site titration assays as described in Chapter III.

Generating mutant complemented TrpSR knockouts

An *E. coli* strain (NM400) competent for recombination, was used to replace the TrpRS gene with a kanamycin resistance cassette in the chromosome of an *E. coli* strain that contains pQE-80L/TrpRS (22). Recombinants were screened by PCR and confirmed by sequencing. Confirmed knockouts were infected with bacteriophage P1 to generate the *TrpS::KAN* P1 lysate (23). Transductants were plated on Evans Blue (EB) to ensure lack of active phage infection;

bacteria with an active phage infection have a lower pH and appear dark green on EB plates, while non-infected cells are pale green (24). The *TrpS::KAN* P1 lysate was used to move the knockout allele into a wild-type *E. coli* (BW27783) that had been previously transformed with an active-site mutant TrpRS on pQE-80L. To increase the rate of evolution, we also introduced pMP6, an arabinose inducible mutagenic plasmid that increases *E. coli*'s basal rate of mutation 322,000-fold (25).

Continuous evolution and fitness assays of mutant complemented TrpRS knockouts

The resulting strain BW27783 *TrpS::KAN* pQE-80L/TrpRS(mutant_X) pMP6 was maintained in continuous liquid culture with antibiotics and inducers for the described amount of time. Every ~24 hours an aliquot of bacteria was removed and mixed with glycerol for storage at -80°C. Periodically, we characterized the fitness of these cultures by comparing their growth rate to the parental strain. Briefly, overnight cultures of the strains to be assayed were diluted to OD_{600} 0.01 and grown at 37°C. Aliquots were removed every 50 minutes for 300 minutes and the OD_{600} measured with a biochrom Ultrospec 10 Cell Density Meter. For each bacterial strain we took the logarithm (base 2) of each timepoint and plotted them versus time to produce a line. The natural logarithm of two was divided by the slope of the line generating the doubling time for each strain. Once an increase in growth rate was achieved we replaced the pQE-80L/TrpRS(mutant_X)^{AmpR} with pQE-80S/TrpRS(mutant_X+1)^{StrepR}, the next mutant in the series of four, via plasmid curing as described in (26).

Results

A chimeric TrpRS urzyme has increased solubility but no detectable catalytic activity

The EKK N-terminal extension greatly increased solubility (~98% soluble) (Figure 4B). However, when combined with the salt-bridge inducing mutations the solubility decreased

(~50% soluble), but the protein expression increased roughly10-fold (data not shown). Despite the increase in solubility and expression of these mutants, we were unable to detect catalytic activity (Figure 4C).

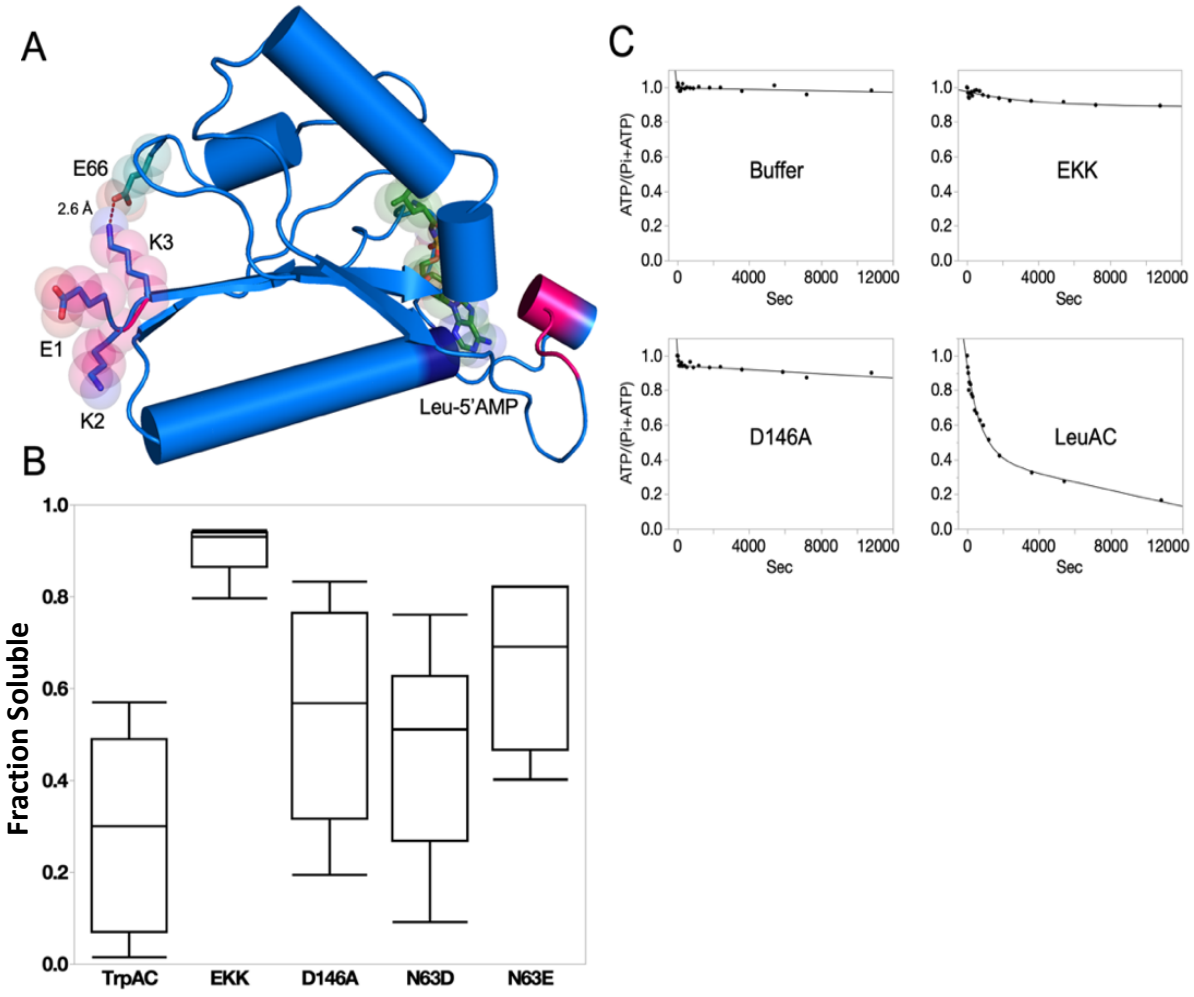


Figure 4. Strategy and implementation of a chimeric TrpAC engineered to increase solubility. A. LeuAC contains an N-terminal charge network increasing its solubility and, possibly, its stability. Integration of this network into TrpAC appeared to be an obvious strategy, including mutating TrpAC N63 to either glutamate or aspartate. The active-site is highlighted by leu-5'AMP (green spheres), and the HVGH (dark blue) and KMSKS (hot pink) sequences. B. Box plot showing the distributions of soluble fraction for different variants. C. Fraction of ATP consumed over time in single-turnover active-site titrations for a buffer control, EKK-TrpAC, EKK-TrpAC(D146A), TEV cleaved MBP LeuAC. The chimeric TrpAC variants are not active.

A TrpRS active-site mutant supports wild-type growth rate

Since we had little knowledge as to the level of TrpRS activity that is required for bacterial survival, we first attempted to generate the TrpRS knockout in the presence of each of the four active-site mutants. We were successful at generating the BW27783 *TrpS::KAN* pQE-80L/TrpRS(K111A) strain, eliminating the need for evolution in the first strain, TrpRS(K195A). Next, we transformed the mutagenic plasmid, pMP6, into the TrpRS(K111A) strain. We determined the growth rate of the parental knockout, an intermediate generation – after plating on EB but before transformation with pMP6, and three different generations after the introduction and induction of pMP6 (day 1, D1; day 37, D37; day 40, D40) (Figure 5). Comparing the doubling time of these strains demonstrates 1) knocking out the native TrpRS and complementing *in trans* with the wild-type TrpRS results in an increased doubling time, 2) TrpRS knockouts complemented with the K111A active-site TrpRS mutant show a significantly increased doubling time as compared to the wild-type complemented knockout, 3) decreased doubling times were observed after growth on Evans Blue for the K111A mutants, 4) a further decrease in the doubling time of K111A was observed after transformation and induction of the pMP6 plasmid.

By day 40 the K111A mutant doubling time was back to the wild-type rate. The pQE-80L/TrpRS(K111A) plasmid was extracted from the strain after day 40 and sequenced. No mutations were found in the coding region or promoter of the TrpRS the gene; therefore, the reduced growth rate is not due to reversion of the TrpRS gene to the wild-type sequence. We attempted plasmid curing to replace the K111A mutant with the K192A mutant. Plasmid curing appeared to have been successful after several weeks of growing in streptomycin to select for pQE-80S/TrpRS(K192A) and checking for the loss of ampicillin resistance conferred by pQE-

80L/TrpRS(K111A). However, upon plasmid extraction and sequencing, we found that the strain had generated a novel plasmid which kept both the streptomycin resistance gene from pQE-80S/TrpRS(K192A) and the more active TrpRS(K111A) synthetase from pQE-80L/TrpRS(K111A). The sequencing was repeated to confirm the unexpected result. To prevent the recombination of multiple plasmids in the future, we went back to the adapted pQE-80L/TrpRS(K111A) strain and knocked out *recA*, a single-stranded DNA binding protein, that is essential for homologous recombination (27). Once the knockout was confirmed by sequencing, we attempted plasmid curing again to replace the pQE-80L/TrpRS(K111A) plasmid with pQE-80S/TrpRS(K192A). During the curing process we were concurrently working on the mutant EKK-TrpAC active-site titrations assays. Upon determining the mutant TrpRS urzymes lacked activity and with many other experiments demanding our attention, we chose to terminate this project.

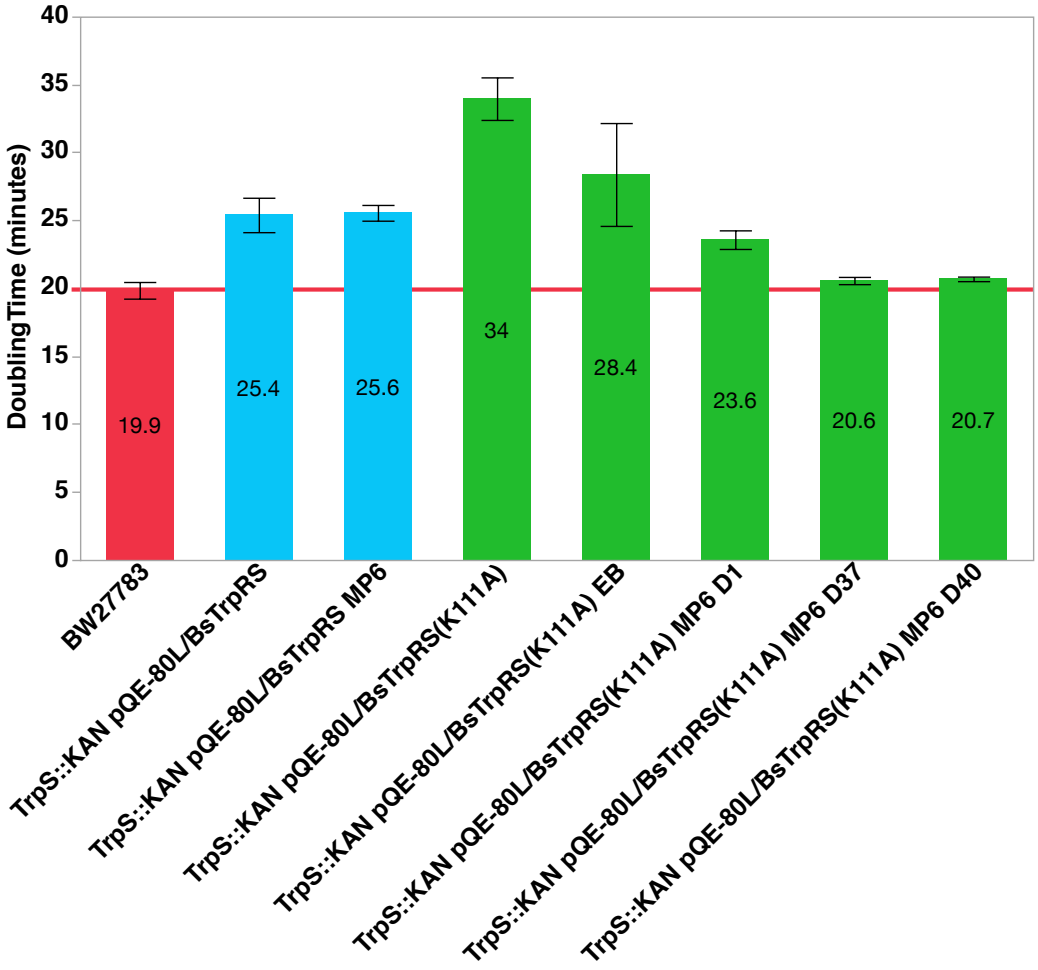


Figure 5. *in vitro* evolution of *E. coli* TrpRS(K111A) mutant. Red bar: wild-type. Blue bars: controls after introduction of plasmid expressing wild-type TrpRS (pQE-80L/BsTrpRS) and knocking out chromosomal copy (TrpS::KAN) and introduction of a mutagenic plasmid (MP6). Green bars: TrpRS knockout complemented with the K111A mutant, after growth on Evans Blue (EB), day 1 (D1) after introduction of pMP6, day 37 (D37), and day 40 (D40). The average doubling time (minutes) of three replicates is shown on bars. A red reference line highlights the average doubling time for the wild-type across the chart.

Discussion

We sought to increase TrpAC solubility and activity by two complementary methods. We were successful at significantly increasing the solubility of a hepta-histidine tagged TrpAC with the addition of three N-terminal charged residues (EKK-TrpAC) and moderately increased the solubility of additional variants (EKK-TrpAC(D146A)/(N63D/E)). However, none of these mutant constructs demonstrate appreciable activity in single-turnover assays.

The evolution of *E. coli* harboring mutant TrpRSs did lead to the derivation of a novel *E. coli* strain that can thrive with an active-site mutant TrpRS. Despite our success in the initial stages of evolution, we were not able to switch out the K111A mutant for the K192A mutant with the initial strategy for plasmid curing. While executing our redesigned plasmid curing strategy, we chose to terminate the project since the mutant TrpAC urzyme constructs, which were to be the final ones introduced into the system, did not display appreciable activity.

Though the TrpAC solubility and evolution project was terminated, there is the potential for this project to be continued, starting with the redesign of TrpAC using Rosetta. The mutant constructs we generated were not evaluated for the effect of the mutations on protein folding. It is likely that the addition of the N-terminal charged residues resulted in TrpAC adopting a new and catalytically inactive conformation. The use of Rosetta to redesign TrpAC could prove to be fruitful and the evolution project could again become relevant. We learned much from this project that will be applied to similar endeavors in the future.

CHAPTER III: Characterization of LeuRS Urzyme Catalytic Activities

The proposal of Rodin and Ohno that aaRSs evolved from a single bidirectional gene (14) motivated us to study the origin of aminoacyl-tRNA synthetases (aaRS) by deconstructing genes of full-length Class I and II aaRS and characterizing the biochemical properties of a hierarchical set of constructs containing their active-site (15-19). Our previous work demonstrated activity for both the smallest constructs, 46 amino acid TrpRS and HisRS peptides (protozymes, (19)), and larger, ~130 amino acid constructs (urzymes, (15-18)), excerpted from full-length TrpRS and HisRS proteins.

No similar studies have been published by other groups until recently, when Koji Tamura's group at Tokyo University of Science (28) re-examined the catalysis we reported from protozymes encoded by opposite strands of a designed gene (19). As we had described, in their hands the Class I protozyme catalyzed amino acid activation by ATP significantly more than maltose-binding protein (MBP) alone when incubated with various amino acids, while the Class II protozyme's activity was closer to, but still greater than, that of MBP alone. Furthermore, both protozymes exhibited greater amino acid activation activity in the absence of any amino acid than did MBP alone. That promiscuous activity, and results presented here, suggest that the aaRS protozymes and urzymes may catalyze other, non-canonical phosphoryl-transfer reactions. Validation of our protozyme results underscore the relevance of our subsequent work with urzymes.

Difficulties preparing high urzyme concentrations for NMR studies (29) by TEV cleavage of the MBP tag and subsequent purification, emphasize the importance of further characterizing urzyme catalytic activities. The tryptophanyl-tRNA synthetase urzyme (TrpAC) had previously been purified as an MBP fusion protein from inclusion bodies and renatured after TEV cleavage (15,16). It has proven difficult to recover active urzyme from that refolding step (unpubished data Tishan Williams and Zhijie Li). Labeling TrpAC with a poly-histidine tag, rather than MBP, also produced an insoluble product. We first described a promising, but ultimately unsuccessful chimeric TrpRS urzyme construct that is significantly more soluble, but inactive. We now describe a LeuAC urzyme, excerpted from *Pyrococcus horikoshii* LeuRS, that is active and soluble under specific buffer conditions, but that cannot be highly concentrated when separated from MBP.

Four aspects of the leucyl-tRNA synthetase system obstructed comprehensive characterization of the kinetic metrics of the LeuAC urzyme: (i) purified LeuRS and LeuAC both retain near stoichiometric amounts of bound leucyl-5'AMP, complicating straightforward interpretation of the [leucine]-dependence of pyrophosphate exchange assays; (ii) it is difficult to prepare cognate $tRNA^{Leu}$ with > 30% acylatability (30-34), limiting confidence in kinetic parameters for aminoacylation by LeuAC; (iii) active-site titrations of LeuRS and LeuAC using α-labeled [^{32}P] ATP both reveal production of the non-canonical nucleotide [^{32}P] ADP, posing significant mechanistic puzzles; (iv) LeuAC and MBP cannot be separated cleanly following TEV cleavage because the urzyme precipitates at moderate concentrations.

We show here that LeuAC fulfills four conditions that support attribution of relevant catalytic activities: significant burst sizes in single-turnover experiments, pyrophosphate exchange activity, aminoacylation of $tRNA^{Leu}$, and binding of significant amounts of [^{14}C]

leucine – presumably as leucyl-5'AMP. Moreover, we describe the catalytic production of ADP, a noncanonical product first observed in a rarely cited study of phosphoryl-transfer reactions exhibited by ArgRS, ValRS, PheRS, and AspRS (35). For these reasons, the experiments described here are essential to guide further work in this challenging area.

Experimental Procedures

Expression and purification of LeuRS and LeuAC

The gene for *Pyrococcus horikoshii* (Ph) LeuRS was synthesized by Gene Universal and expressed from pET-11a in BL21-CodonPlus (DE3)-RIPL (Agilent). Cells were grown at 37°C and induced with 300 μM IPTG for 4 hours then harvested and stored overnight at -20°C. The cell pellet was resuspended in 1x Ni-NTA buffer (20 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 mM β-ME) plus cOmplete protease inhibitor (Roche) and lysed by three 15K psi passes on an Avestin Emulsiflex c5. Cell debris was pelleted at 4°C for 30 minutes at 20K rpm. The soluble fraction was heated at 80°C for 30 minutes to denature native *Escherichia coli* proteins. The heated cell extract was then pelleted, and the soluble material was loaded on to an equilibrated Ni-NTA column. The column was washed with three volumes 1x Ni-NTA buffer, then protein was eluted in a stepwise fashion with imidazole concentrations of 40, 80, 100, 200, and 500 mM imidazole. The fractions containing the protein of interest were pooled and dialyzed overnight against 200 mM HEPES, pH 7.4, 450 mM NaCl, 100 mM KCl, 10 mM β-ME. The following day the dialyzed protein was concentrated and mix to 50% (V/V) glycerol and stored at -20°C.

LeuAC was expressed as an MBP fusion from pMAL-c2x in BL21Star(DE3) (Invitrogen). Cells were grown, induced, harvested, and lysed similarly to Ph LeuRS with the distinct difference of being resuspended in Optimal Buffer (20 mM Tris, pH 7.4, 1 mM EDTA, 5

mM β-ME, 17.5% Glycerol, 0.1% NP40, 33 mM $(NH_4)_2SO_4$, 1.25% Glycine, 300 mM Guanidine Hydrochloride) plus cOmplete protease inhibitor (Roche). LeuAC crude extract was then pelleted at 4°C for 30 minutes at 15K rpm to remove insoluble material. The extract was diluted 1:4 with Optimal Buffer and loaded onto equilibrated Amylose FF resin (Cytiva). The resin was washed with five column volumes of buffer and the protein was eluted with 10 mM maltose in Optimal Buffer. Fractions containing the protein of interest were concentrated and mixed to 50% glycerol and stored at -20°C. All protein concentrations were determined using the Pierce™ Detergent-Compatible Bradford Assay Kit (Thermo Scientific). Experimental assays were performed either with the intact MBP-LeuAC fusion protein or with TEV cleaved LeuAC that was not separated from MBP. Purity and cleavage efficiency was determined by running samples on PROTEAN® TGX (Bio-RAD) gels. Some experiments used a second LeuAC variant, from a more recent Rosetta design algorithm in which we attempted to modify surface residues to increase solubility (Matt Cummins, unpublished). Amino acid sequences for both variants are given in the Appendix 1 data and compared to the native Ph LeuRS sequence. We could not detect any significant differences between results from the two different LeuAC fusion constructs.

Nitrocellulose filter binding of the LeuRS/LeuAC leucyl-adenylate complex

1 μM LeuRS or LeuAC was mixed with 25 μM [^{14}C] leucine in 50 mM HEPES, pH7.5, 20 mM $MgCl_2$, 50 μM ATP, and inorganic pyrophosphatase. A zero time point was collected prior to addition of the enzyme. Reactions were incubated at 37°C and aliquots were removed at indicated time points and spotted onto prewashed nitrocellulose filters. The filters were washed with reaction buffer lacking ATP and leucine and allowed to dry before scintillation counting in 5 ml of BetaMax on a Beckman Packard.

Size exclusion chromatography of the LeuAC leucyl-adenylate complex

1-8 μM of LeuAC was incubated at 37°C for 36 minutes in buffer containing 50 mM HEPES, pH 7.5, 20 mM $MgCl_2$, 4 mM ATP, and 40 μM [^{14}C] leucine or 5 mM leucine with comparable cpm of [^{14}C] leucine. 2 ml of Sephadex G-15 resin (Cytiva) was loaded into a disposable gravity flow column (120 x 10 mm) and equilibrated with five column volumes of 50 mM HEPES, pH 7.5 and 20 mM $MgCl_2$. 100 μl reactions were applied to the column and eluted with 3 ml of buffer. Single-drop fractions (~40 μl) were collected and analysed either by UV-Vis (Nanodrop) or scintillation counting (Beckman Packard). Molar extinction coefficients of the MBP-LeuAC fusion protein (9.63e4) and ATP (2.31e3) are in the ratio 41.7:1, so ATP absorbance of a stoichiometric ATP:LeuAC complex is at most 2.4% of the total, and therefore essentially insignificant. The obvious separation of LeuAC from lower molecular weight compounds enabled curve fitting to an exponentially-modified normal distribution (18) with R^2 values >0.97 using JMP™ Pro. Quantitative estimation of the relative volumes of overlapping peaks was done using the eluted volume as the Y column and the corresponding value as the frequency in the JMP™ Pro distribution module and fitting to a continuous 3-normal function.

Pyrophosphate (PPi) exchange assays

PPi exchange assays were performed as described (9,36), with some variation when we observed inhibition by KF. Enzyme concentrations were generally 3 μM (LeuAC) and 0.3 μM (*E. coli* LeuRS) in 50 μL assay volumes. [^{32}P] ATP formed at time intervals was separated from [^{32}P] PPi by thin layer chromatography on polyethyleneimine plates (Scientific Adsorbents) that had been pre-run in water and developed using a mobile phase of 750 mm KH_2PO_4, pH 3.5 and 4 M urea, at a running temperature of 25°C. Radioactive spots were detected by phosphorimaging on a Typhoon Scanner (Cytiva) and quantified using the ImageJ measure function (NIH).

Single-turnover active-site titration assays

3 μM of protein was added to 1x reaction mix (50 mM HEPES, pH 7.5, 10 mM $MgCl_2$, 10 μM ATP, 50 mM amino acid, 1 mM DTT, inorganic pyrophosphatase, and either α- or γ-labeled [^{32}P] ATP) to start the reaction. Timepoints were quenched in 0.4 M sodium acetate 0.1% SDS and kept on ice until all points had been collected. Quenched samples were spotted on TLC plates, developed in 850 mM Tris, pH 8.0, dried and then exposed for varying amounts of time to a phosphor image screen and visualized with a Typhoon Scanner (Cytiva). The ImageJ measure function was used to quantitate intensities of each nucleotide. The time-dependent of loss (ATP) or de novo appearance (ADP, AMP) of the three adenine nucleotide phosphates were fitted using the nonlinear regression module of JMP™ Pro to equation (1):

$$(1) \qquad Product_{calc} = A^{(-kchem*seconds)} + - k_{cat}*seconds + C$$

where k_{chem} is the first-order rate constant, k_{cat} is the rate of turnover, A is the amplitude of the first-order process, and C is an offset.

$tRNA^{Leu}$ aminoacylation assays

A plasmid encoding the *P. horikoshii* $tRNA^{Leu}$ (TAG codon) was synthesized by Integrated DNA Technologies and used as template for PCR amplification of the tRNA and upstream T7 promoter and downstream Hepatitis Delta Virus (HDV) ribozyme. The PCR product was used directly as template for T7 transcription. Following a 4-hour transcription at 37°C the RNA was cycled five times (90°C for 1 min, 60°C for 2 min, 25°C for 2 min) to increase the cleavage by HDV. The tRNA was purified by urea PAGE and crush and soak extraction. The tRNA 2'-3' cyclic phosphate was removed by treatment with T4 PNK (New

England Biolabs) following the manufacturer's protocol. The tRNA was then phenol chloroform isoamyl alcohol extracted, filter concentrated, aliquoted, and stored at -20°C.

Aminoacylation by *P. horikoshii* LeuRS followed incorporation of [^{14}C] leucine into acid precipitable material. tRNA was folded by heating in 30 mM HEPES, pH 7.5, 30 mM KCl to 90°C for 2 minutes. The tRNA was then cooled linearly (drop 1°C/30 seconds) until it reached 80°C when $MgCl_2$ was added to a final concentration of 10 mM. The tRNA continued to cool linearly until it reached 20°C. Enzyme was added to 0.95 μM (for MM experiments; see Appendix 5 for concentration in other experiments) and aliquots were taken at indicated timepoints (8 minutes for MM experiments) and spotted onto 10% TCA prewashed Whatman 3MM filters, washed three times with 5 ml cold 5% TCA and a final wash of cold ethanol. Dried filters were counted in 5ml BetaMax (MP BioMedicals) on a Beckman Packard scintillation counter.

[^{14}C] leucine proved too insensitive for accurately measuring acylation by the weaker LeuAC catalyst. LeuAC aminoacylations were performed in 50 mM HEPES, pH 7.5, 10 mM $MgCl_2$, 20 mM KCl, 5 mM DTT with indicated amounts of ATP and amino acids. Desired amounts of unlabeled tRNA were mixed with [α^{32}P] labeled tRNA and folded similarly to that used for Ph LeuRS acylations. Re-folded tRNA was mixed with buffer and then a zero timepoint collected prior to initiation of the reaction by addition of the enzyme to 1.2 μM (for MM experiments; see Appendix 5 for concentration in other experiments). Indicated timepoints (five mintues for MM experiments) were quenched by adding into a solution of 0.4 M sodium acetate, pH 5.2, 6.25 mM Zn Acetate, 10U P1 nuclease and stored on ice until all time points had been collected. Quenched samples were incubated at 37°C for 10 minutes to allow digestion of the

tRNA by the P1 nuclease. Samples were spotted on pre-run TLC plates and developed in 10% NH_4Cl, 5% acetic acid.

Dried TLC plates were exposed overnight to a phosphor screen and visualized on a Typhoon Scanner. ImgaeJ measure function was used to obtain intensities for AMP and leucyl-AMP. tRNA concentrations were corrected for fraction acylatable tRNA as estimated from extended time course acylations with full-length Ph LeuRS. For single-turnover analysis, loss of unacylated [active] tRNA was fitted to equation (1).

Data processing and statistical analysis

Phosphorimaging screens of TLC plates were densitometered using ImageJ. Data were transferred to JMP™ Pro 16 via Microsoft Excel (version 16.49), after intermediate calculations. The nonlinear fitting module was used to fit all active-site titration and Michaelis-Menten datasets. Factorial design matrices in Tables 1, 2 were processed using the Fit model multiple regression analysis module of JMP™ Pro, using an appropriate form of equation (2)

$$(2) \qquad Y_{obs} = \beta_0 + \Sigma\beta_i * P_i + \Sigma\beta_{ij} * P_i * P_j + \varepsilon$$

where Y_{obs} is a dependent variable, usually an experimental observation, β_0 is a constant derived from the average value of Y_{obs}, β_i and β_{ij} are coefficients to be fitted, $P_{i,j}$ are independent variables used as predictors, and ε is a residual to be minimized. All rates were converted to free energies of activation, $\Delta G^{\ddagger} = -- RT\ln(k)$, before regression analysis because free energies are additive, whereas rates are multiplicative. For example, the activation free energy for the first-order decay rate in single-turnover experiments is $\Delta G^{\ddagger} k_{chem}$. It is worth emphasizing that multiple regression analyses of factorial designs exploit the replication inherent in the full

collection of experiments to estimate the experimental variance based on t-test P-values, in contrast to the presenting error bars showing the variance of individual datapoints. In some cases, multiple regression analyses reported here also entail experimental replicates, which enhance the associated analysis of variance.

Results

TEV cleaved MBP-LeuAC fusion protein is pure enough to minimize contaminating activities

TEV cleavage of the MBP-LeuAC fusion protein does not go to completion (Figure 6A). Estimates derived from band intensities (Figure 6B) using tryptophan residues in each relevant species (MBP-LeuAC fusion, 10; LeuAC, 3; MBP 7; *E. coli* LeuRS, 24) suggest that TEV cleavage leaves 14% of the fusion protein, and that the maximum amount of *E. coli* LeuRS would be at most 1 in 10^3. Estimates of the active fraction of LeuAC molecules (0.2-0.4) evident in Table 1 are >200 fold greater than this value, which thus cannot account for any catalytic activities documented here.

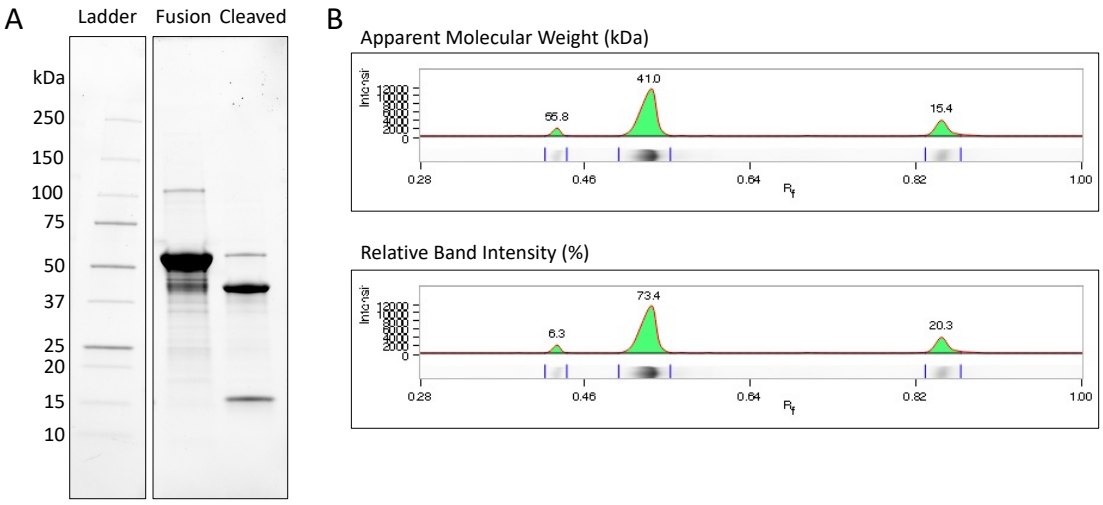


Figure 6. Purity of LeuAC urzyme. A. PROTEAN® TGX PAGE gel of purified LeuAC-MBP fusion protein and its TEV cleaved products. Visualization is proportional to the tryptophan content of each band, as noted in the text. B. Densitometric scan of the gel in A. Top panel shows apparent molecular weight of detected peaks (MBP-LeuAC 55.8, MBP 41.0, and TEV cleaved LeuAC 15.4 kDa) and bottom panel shows relative band intensities.

Active-site titration assays with [^{32}P] ATP

The urzymes we have studied all bind tightly to the aminoacyl-5'AMP intermediate product, so that product release is rate-limiting in single-turnover enzymatic assays with substrate-level amounts of enzyme (9,36). Under these conditions, the amplitude of the first-order process, or burst size, can be used to estimate the number of catalytically active molecules in the added enzyme (36). [γ^{32}P] ATP is rapidly lost at different rates over a ~90 minute time-course in the presence of either full-length LeuRS, MBP-LeuAC fusion protein, or TEV cleaved LeuAC. With [α^{32}P] ATP LeuRS produced not only the expected [^{32}P] AMP, but also [^{32}P] ADP (Figure 3A), which LeuAC produced in near stoichiometric quantities (Figure 7B). The high amplitude of this unexpected, non-canonical product helps to validate the authenticity of LeuAC catalytic functionality but poses significant mechanistic questions.

Routine use of AST assays—to characterize preparations of MBP-LeuAC fusion proteins, and in searching for high-affinity inhibitors—produced highly replicated data (Table 1), from which we can compare behaviors of the fusion protein and its TEV-cleaved products. Qualitatively, TEV cleaved LeuAC produces higher first-order rate constants, and larger burst sizes for the loss of ATP and the appearance of the two other adenine nucleotides than does the intact fusion protein (Figure 7C). Replicated AST assays with [α^{32}P] ATP (Table 1) verified the high statistical significance of these observations. Two predictors, TEV cleavage and AMP, explain 93% of the variance in activation free energies for the first-order rates, $\Delta G^{\ddagger}k_{chem}$, with t-test P values of ~2 x 10^{-10} and 2 x 10^{-7} (Figure 8A).

TEV cleavage reduces $\Delta G^{\ddagger}k_{chem}$ for all three reactions by ~0.5 kcal/mole, and first-order rate of AMP production is further increased by a similar amount from that observed for the loss of ATP and appearance of ADP. Consequently, the initial appearance of AMP occurs with a rate

slightly greater than the loss of ATP or the appearance of ADP (Figure 8A). The dependence of the amplitude parameter, A, representing the pre-steady-state burst size, is also significantly suppressed in the MBP fusion protein (Figure 8B) relative to that of TEV cleaved LeuAC. The larger burst sizes, together with the significant functional modification in the fusion protein, strongly and independently imply that the time courses in Figure 7C cannot be attributed to contaminating catalysts, and therefore must be considered functionalities of LeuAC itself.

Near stoichiometric appearance of the non-canonical product, ADP, is unexpected considering the consensus mechanisms of amino acid activation by aaRS, which proceeds by the concomitant release and hydrolysis of PPi by the pyrophosphatase present in the assay mix, but which cannot be traced if the [^{32}P] label is in the γ position. It recalls a generally overlooked early publication from Zamecnik's laboratory (35) that some aaRSs also catalyze ADP production. ADP production by the full-length enzyme represents a small fraction (~ 3 %) of the total active-site concentration (Figure 7A). The additional protein mass of the full-length LeuRS, in comparison with LeuAC, apparently reduces the production of ADP. We consider possible implications of the near-stoichiometric ADP production by LeuAC further in the Discussion.

Sample	ATP	ADP	AMP	TEV	k_{chem}	$\Delta G^{\ddagger}$ (k_{chem})	k_{cat}	$\Delta G^{\ddagger}$ (k_{cat})	C	A
MBP_LeuAC2	1	0	0	0	0.00166	3.79	2.66E-05	6.24	0.6	0.2
MBP_LeuAC2, LSA	1	0	0	0	0.00172	3.77	2.21E-05	6.35	0.49	0.3
MBP_LeuAC2, APQ	1	0	0	0	0.00181	3.74	1.65E-05	6.52	0.48	0.3
MBP_LeuAC2	0	1	0	0	0.00147	3.86	2.27E-05	6.33	0.37	0.2
MBP_LeuAC2, Vanadate	0	1	0	0	0.00153	3.84	1.15E-05	6.73	0.5	0.3
MBP_LeuAC2, LSA	0	1	0	0	0.00177	3.75	1.99E-05	6.41	0.48	0.3
MBP_LeuAC2, APQ	0	1	0	0	0.00168	3.78	1.38E-05	6.63	0.5	0.3
MBP_LeuAC2	0	0	1	0	0.00268	3.51	8.22E-07	8.29	0.04	0
MBP_LeuAC2_Vanadate	0	0	1	0	0.0034	3.36	2.78E-06	7.57	0.03	0
MBP_LeuAC1	1	0	0	0	0.00131	3.93	3.55E-06	7.43	0.8	0.2
LeuAC1_Tev_cleaved	1	0	0	1	0.00326	3.39	0.00002	6.41	0.27	0.6
LeuAC1_Tev_cleaved	1	0	0	1	0.00321	3.40	2.07E-05	6.38	0.26	0.6
LeuAC1_Tev_cleaved	0	1	0	1	0.00312	3.42	3.16E-06	7.50	0.71	0.6
LeuAC1_Tev_cleaved	0	1	0	1	0.00316	3.41	2.1E-06	7.74	0.69	0.6
LeuAC1_Tev_cleaved	0	0	1	1	0.00494	3.14	1.59E-05	6.54	0.04	0
LeuAC1_Tev_cleaved	0	0	1	1	0.00767	2.88	1.92E-05	6.43	0.03	0
LeuAC1_Tev-cleaved, Ile	1	0	0	1	0.00378	3.30	0.00003	6.16	0.36	0.5
LeuAC1_Tev-cleaved, no aa	1	0	0	1	0.0037	3.31	0.00003	6.16	0.34	0.5
LeuAC1_Tev-cleaved, Ile	0	1	0	1	0.00393	3.28	0.00001	7.06	0.6	0.5
LeuAC1_Tev-cleaved, no aa	0	1	0	1	0.00387	3.29	0.00001	6.88	0.62	0.5

Table 1. **Design matrix for dependencies of [^{32}P] ATP active-site titrations**. Independent variables are [ATP], [ADP], [AMP], TEV cleavage. Dependent variables– k_{chem}, kcat, A, C, and associated free energies–are from fitting to Eqn (1). LeuAC1 and LeuAC2 are labelled according to Appendix 1.

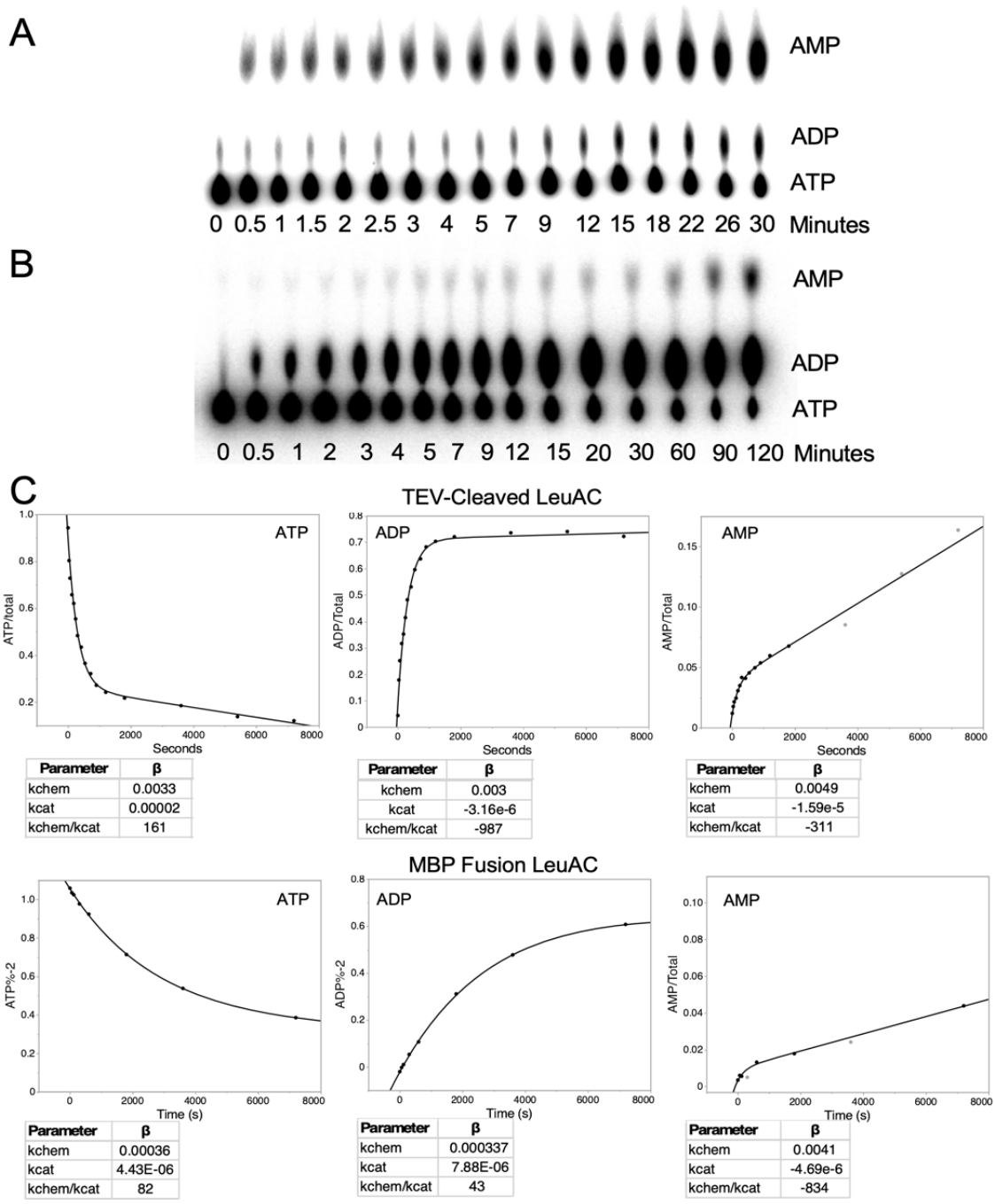
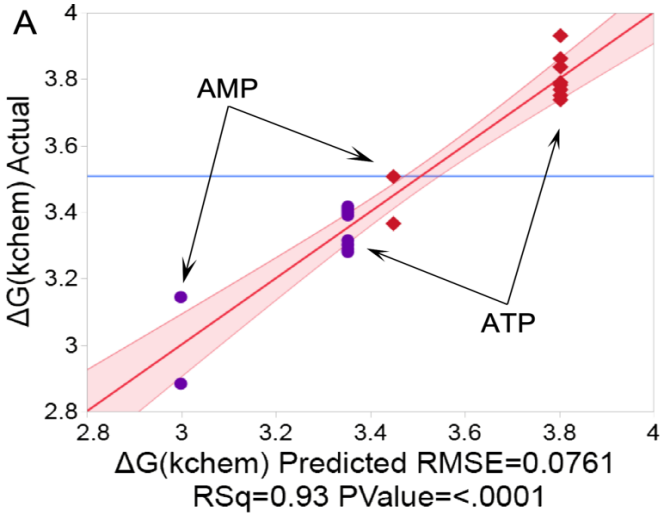
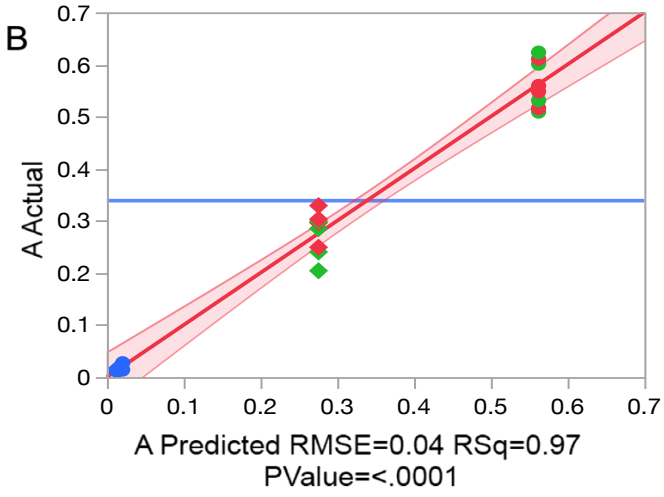


Figure 7. Time-courses for the appearance of the three adenine nucleotides in active-site titration experiments using [α-^{32}P] ATP. Thin-layer chromatograms for full-length *P. horikoshii* LeuRS (A) and TEV-cleaved MBP LeuAC (B) performed using [α-^{32}P] ATP. Assays were performed as described in the text. Note the difference in time scales between the full-length and urzyme proteins, as well as the different proportions of the appearance of [^{32}P] ADP and [^{32}P] AMP. C Fitted plots of representative selections from experiments in Table 1. Plots done by both co-authors over a span of 4 years with different enzyme preparations are accompanied by fitted β coefficients and standard errors and show qualitative differences induced by TEV cleavage of the LeuAC-MBP fusion protein.



Term	β, kcal/mole	Std Error	t Ratio	Prob>|t|
Intercept	3.803	0.026	149.07	<.0001*
AMP	-0.353	0.043	-8.31	2.16E-07
TEV-cleaved	-0.450	0.034	-13.23	2.20E-10



Term	β	Std Error	t Ratio	Prob>|t|
Intercept	0.30	0.013	22.63	<.0001*
AMP	-0.40	0.022	-17.99	1.0E-11
TEV-cleaved	0.23	0.018	12.9	7.4E-10
(AMP)*(TEV)	-0.28	0.045	-6.24	1.2E-05

Figure 8. Multiple regression analyses of the dependence of $\Delta G^{\ddagger}k_{chem}$ and burst size on TEV cleavage and nucleotide. A. Dependence of $\Delta G^{\ddagger}k_{chem}$ on nucleotide and TEC cleavage. Circles represent TEV-cleaved and diamonds represent the MBP fusion protein. B. Analysis of dependence of burst size 'A' on nucleotide and TEV cleavage. Data drawn from AST experiments in Table 1 and performed using both [α-^{32}P] and [γ-^{32}P] ATP. Circles represent TEV cleaved, diamonds fusion protein, blue and green amplitude calculated from AMP and ADP production, respectively, and red amplitude calculated from ATP loss. R^2 and overall F-ratio P-values are included in the X axis labels. Tables include β coefficients with standard deviations, Student t-tests, and P values.

Size exclusion chromatography of LeuAC demonstrates a significant fraction of active protein

The curious development of ADP in active-site titrations, discussed above, underscored the need for an orthogonal measurement of the active fraction of enzyme in preparations of both full-length *P. horikoshii* LeuRS and LeuAC derived from it. For this purpose, we first measured the retention of [^{14}C] leucine bound to LeuRS and LeuAC collected on nitrocellulose filters (36). Although LeuRS retained 0.48 ± 0.13 moles of [^{14}C] leucine/mole enzyme, LeuAC was much less reproducible, retaining 0.10 ± 0.07 moles/mole enzyme. For that reason, we also used size exclusion chromatography on Sephadex G-15 to separate TEV cleaved LeuAC from low molecular weight substrates.

Sephadex G-15 size exclusion profiles of LeuAC, ATP, and reaction mixtures with and without additional leucine at 5 mM were recorded at 280 nm (Figure 9A) and that of ATP was also recorded at 260 nm. Although the stoichiometric excess of ATP in the reaction mixtures contributed significantly to the absorbance profile, the molar extinction coefficients of ATP (2.31E3) and MBP-LeuAC (9.63E4) are in the ratio = 0.024. Thus, ATP bound to LeuAC is barely detectable at 280 nm, allowing quantitation of both eluted products. Integrated A280 of the two reaction mixtures (26.1 and 28.2) are both within experimental error of the total A280 (27.6) of the LeuAC (7.8) and ATP (19.8).

Total CPM from [^{14}C] leucine and A280 absorbance profiles for both reactions were scaled together to evaluate difference profiles (Figure 9B). Quantitative deconvolution of peak integrals, described in the Methods section, demonstrate the mutual influence of LeuAC and leucine, which shift each other's elution toward the excluded volume. Addition of 5 mM leucine to the reaction mixture shifts 4.9 μM of the 8.52 μM LeuAC to the left (green profile); whereas addition of 8.52 μM LeuAC shifts 0.14 of the 50 μM or ~7.3 μM of the [^{14}C] leucine to the left

(blue profile). Comparable proportions of leucine and LeuAC are perturbed by the other. Nearly 60% of the LeuAC binds reversibly to exogenous leucine. Inasmuch as the [^{14}C] leucine shift is about a third of the difference between the eluted position of LeuAC and ATP, the half-life of bound leucine is roughly a third of the duration of transit through the column (~18 minutes). Quantitative analysis of this behavior would have required that the elution buffer contain constant ligand concentrations (37) which was impractical owing to the cost of radiolabeled ligand. Qualitative analysis in the Discussion sets lower limits on both stoichiometry and affinity of the active LeuAC•leucyl-5'AMP complex.

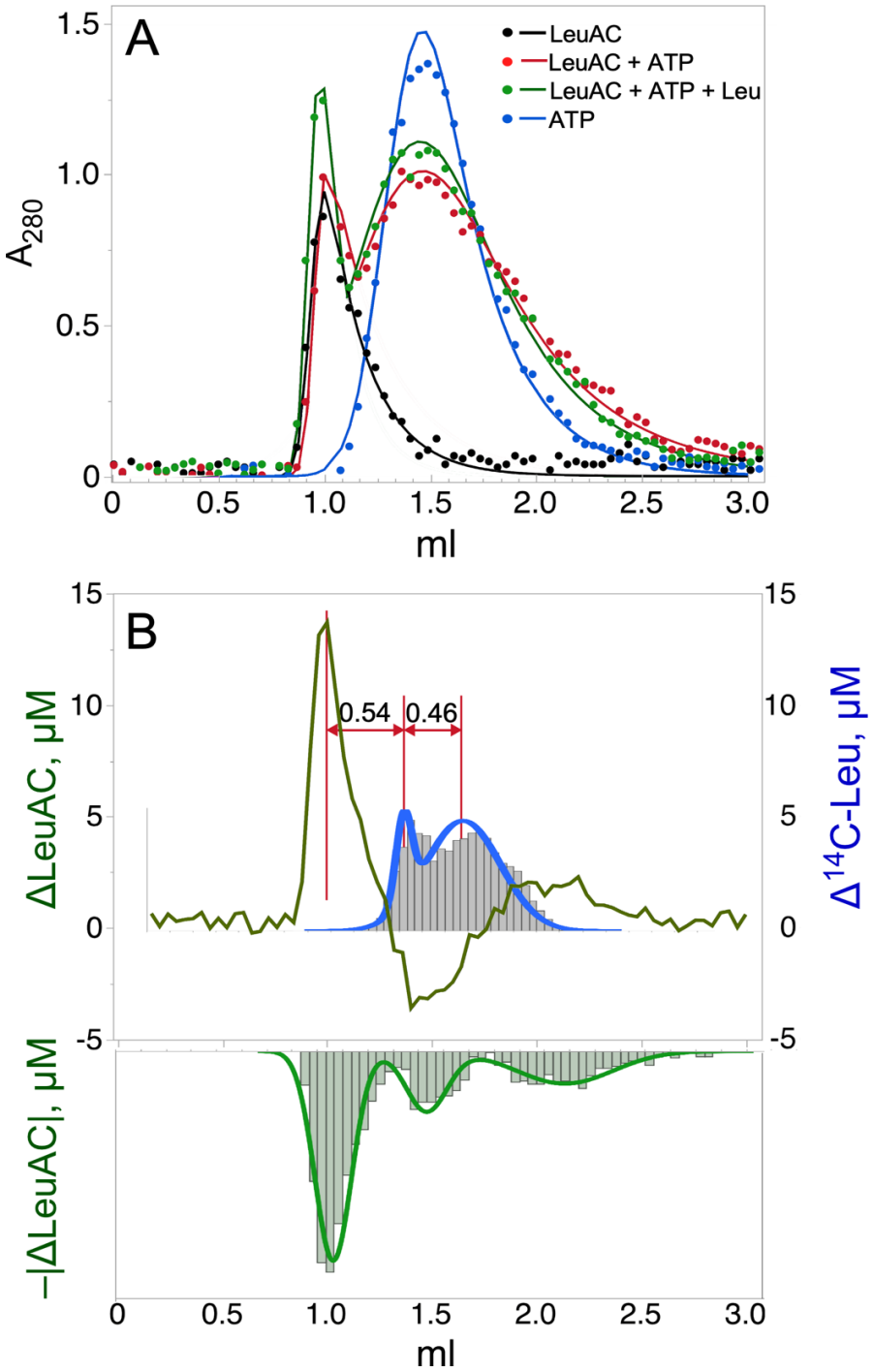


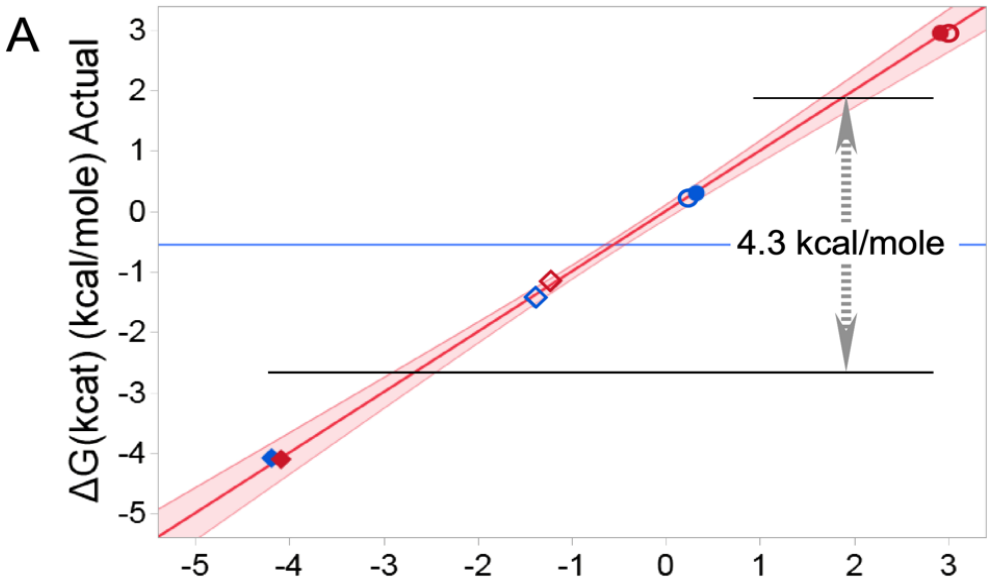
Figure 9. Quantitative size exclusion chromatography of TEV cleaved LeuAC on Sephadex G-15. A. Profiles for LeuAC, ATP, and reaction mixtures either with or without 5mM additional leucine. Experimental points were fitted to a four-parameter exponentially modified normal distribution (38). B. Quantitative deconvolution of overlapping peaks in ±LeuAC and ± leucine difference profiles show the mutual acceleration by LeuAC and substrate leucine of the other's elution (upper panel); negative the absolute value of the change in LeuAC concentration with the added leucine.

The LeuAC MBP fusion protein catalyzes pyrophosphate exchange

We compared PP_i exchange activity of MBP-LeuAC and full-length LeuRS in the two-level, three-factor experiment in Table 2. Regression analysis in Figure 10 shows that β coefficients for the regression model are all statistically significant, with t-test -values < 0.01. Catalysis by LeuRS is, on average, ~1500 times faster than that by LeuAC, owing to the 4.3 kcal/mole difference between mean activation free energies (Figure 10A). Qualitative differences between the LeuRS and LeuAC activities are indicted by the two, 2-way interaction free energies. Potassium fluoride (KF) is used in PPi exchange assays to eliminate activity by contaminating phosphatases (9). LeuAC-catalyzed incorporation of [^{32}P] PPi into ATP is inhibited by KF, whereas that by LeuRS is unaffected (Figure 10B). LeuRS is stimulated by excess leucine, whereas LeuAC is not (Figure 10C). Separate tests for enzymatic hydrolysis of [^{32}P] ATP and [^{32}P] PPi by the MBP-LeuAC did not exceed background. Nonetheless, the experiment summarized in Figure 10 does not authenticate the biologically relevant PP_i exchange activity of LeuAC, because it does not demonstrate [leucine] dependence.

EXPT	LeuAC	Leucine	KF	[Enz], M	Δ[ATP]/time	k_{cat} (/s)	$\Delta G^{\ddagger}_{(kcat)}$ (kcal/mole)
1	1	0	0	3.00E-06	2.020E-06	6.7E-01	0.234
2	1	0	1	3.00E-06	2.059E-08	6.9E-03	2.949
3	1	1	0	3.00E-06	2.004E-06	6.7E-01	0.239
4	1	1	1	3.00E-06	2.073E-08	6.9E-03	2.945
5	-1	0	0	3.00E-07	3.145E-06	1.0E+01	-1.391
6	-1	0	1	3.00E-07	2.281E-06	7.6E+00	-1.201
7	-1	1	0	3.00E-07	3.118E-04	1.0E+03	-4.112
8	-1	1	1	3.00E-07	3.154E-04	1.1E+03	-4.119

Table 2. **Design matrix for dependences of PPi exchange**. Independent variables are LeuAC (1) or LeuRS (-1), [Leucine] (1,0) potassium flouride (0,1), and enzyme concentration. Dependent variables are the rate of ATP formation and derived quantities: kcat = rate/[Enz], $\Delta G^{\ddagger}$ (k_{cat}) = –RT ln(k_{cat}). The LeuAC2 construct was used as the MBP fusion in these experiments.



Term	β, kcal/mole	Std Error	t Ratio	Prob>|t|
Intercept	-0.55	0.042696	-12.95	0.0059*
LeuAC	0.79	0.042696	18.48	0.0029*
Leucine	-1.41	0.049301	-28.59	0.0012*
KF	1.40	0.049301	28.42	0.0012*
LeuAC*KF	1.31	0.049301	26.56	0.0014*
LeuAC*Leu	1.41	0.049301	28.6	0.0012*

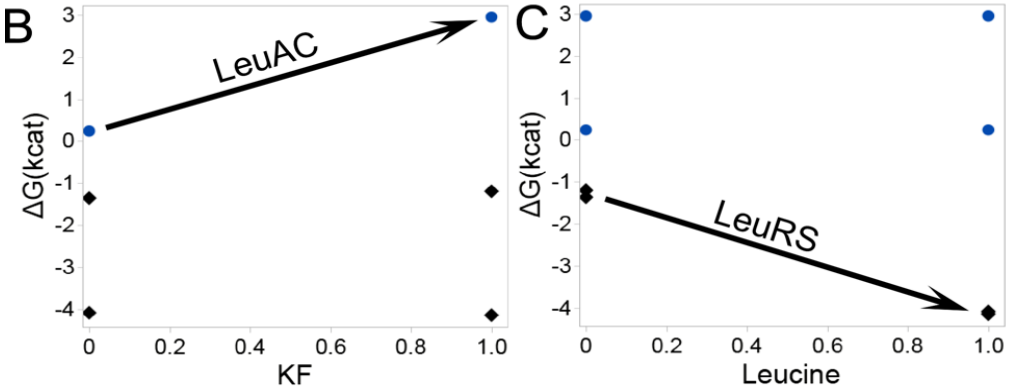


Figure 10. Regression analysis of PP_i exchange activities based on the design matrix in Table 2. A. Multivariate model for the activation free energy, $\Delta G^{\ddagger}k_{cat}$, and table of β coefficients with standard deviations, Student t-tests, and P-value probabilities. Symbols: diamonds = LeuRS, circles = LeuAC, empty symbols = no added leucine, red symbols = plus KF. Supplemental plots of the effects of [leucine] (B) and KF (C) clarify the interpretation of two-way interactions, as noted in the text.

Both LeuRS and LeuAC retain tightly bound leucyl-5'AMP

Amino acid-dependence of PPi exchange is an important criterion for confidence in the authenticity of catalytic activity. Lack of that dependence for LeuAC PPi exchange activities created concern. Previous studies (34) showed that full-length LeuRS purifies with near stoichiometric amounts of bound leu-5'AMP. Those authors incubated full-length *E. coli* LeuRS with excess $tRNA^{Leu}$ to obtain sufficient unliganded LeuRS for studies of the partitioning of pre- and post-transfer editing. That approach was impractical owing to the relatively low stability of LeuAC in 2M urea necessary to dissociate $tRNA^{Leu}$ from the acylation complex.

Extended dialysis of MBP-LeuAC fusion protein against six, successive 1000-fold volumes of 40 mM Tris-HCl buffer, pH = 8.0, 100 mM NaCl, 20 % glycerol, 10 mM BME increased PPi exchange 1.7-fold (P = 0.01). Equilibrium dialysis is inefficient at removing tightly bound ligands because the equilibrium concentration of free ligand inside the membrane is so low that the gradient across the membrane fails to remove much of the bound ligand, even after extensive changes. The dialysis experiment nonetheless implicates the presence of a tightly bound ligand. Nitrocellulose binding and gel filtration (Figure 9) suggest that this ligand is leucyl-5'AMP, which can support [leucine]-independent PPi exchange. Persistence of that ligand through extensive dialysis is consistent with the expected ~1000-fold increase in affinity of the 5'adenylate, relative to the free amino acid, with pre-steady-state bursts observed in single-turnover assays (Figures 7, 8), and with lower limits of affinity from size exclusion chromatography (See Discussion).

LeuRS, MBP-LeuAC fusion, and TEV cleaved LeuAC catalyze tRNA^{Leu} aminoacylation to different extents

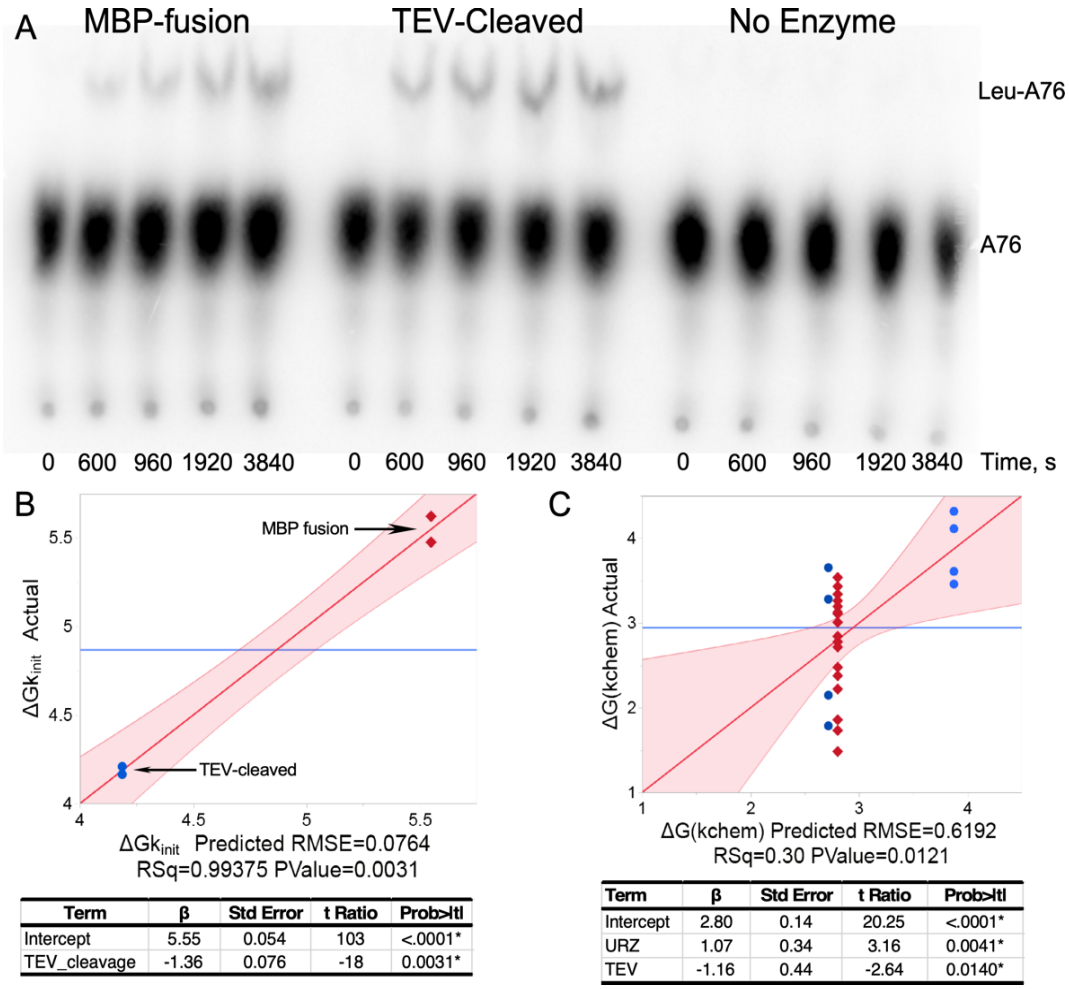
From many perspectives, the key aaRS urzyme catalytic activity is acyl transfer of activated amino acid to cognate tRNA. Here, we demonstrate catalysis of acyl-transfer by LeuAC for the first time (Figure 11A). TEV cleavage increases LeuAC aminoacylation rates by 10-fold (Figure 11B). The biphasic time dependence of aminoacylation is described in the following paragraph. An ensemble of 28 acylation assays (Appendix 5) reveals that the principal determinants of $\Delta G^{\ddagger}(k_{chem})$ are (i) whether the catalyst is full-length LeuRS or LeuAC urzyme and (ii) whether the urzyme is the MBP fusion or is TEV cleaved (Figure 11C). β-coefficients for the overall effect of TEV cleavage (–1.36 kcal/mole; Figure 7B) and its effect on $\Delta G^{\ddagger}k_{chem}$ (–1.16; Figure 11C), suggest that about 85% of the increase derives from enhancement by TEV cleavage of the first-order rate, k_{chem}, rather than from increases in turnover, k_{cat}.

Single-turnover kinetics furnish three metrics for comparing LeuRS, LeuAC fusion protein, and TEV cleaved LeuAC

Curiously, k_{chem} values for LeuRS and TEV cleaved LeuAC are both faster than that of the MBP fusion protein (Figure 11C). That pattern – positive β-coefficients for the urzyme, relative to full-length and negative β-coefficients for TEV cleavage – recurs in several contexts and provides important and orthogonal evidence for the authenticity of LeuAC catalysis.

Time courses for aminoacylation by LeuRS and LeuAC (Figure 12A) both exhibit bi-phasic kinetics that fit with very small unexplained variances (R^2 >0.98) to equation (1) for a first-order decay of a single-turnover and steady state turnover (Figure 12B). The fitting precision, the physical interpretation of β coefficients as the amplitude of the first-order phase, A, and the first-order and steady-state rates, k_{chem} and k_{cat}, and the roughly parallel changes in the

three parameters in amino acid activation and acyl-transfer (Appendix 4 and Figures 7,8,11,12) argue that they are appropriate metrics. Comparison of these metrics for LeuRS, MBP-LeuAC, and TEV cleaved LeuAC are, in turn, evidence for the authenticity of amino acid activation and tRNA acylation by all three catalysts.



Term	β	Std Error	t Ratio	Prob>|t|
Intercept	5.55	0.054	103	<.0001*
TEV_cleavage	-1.36	0.076	-18	0.0031*

Term	β	Std Error	t Ratio	Prob>|t|
Intercept	2.80	0.14	20.25	<.0001*
URZ	1.07	0.34	3.16	0.0041*
TEV	-1.16	0.44	-2.64	0.0140*

Figure 11. Aminoacylation of $tRNA^{Leu}$ by LeuAC increases upon TEV cleavage of the MBP fusion. A TLC of time courses for the MBP fusion, TEV-cleaved LeuAC, and a minus enzyme control. B. Comparison of the initial rates for matched assays similar to that in A on different days, with the corresponding regression analysis indicating a decrease of –1.36 in $\Delta G^{\ddagger}$ for the initial rate, corresponding to an ~10-fold increase in rate. C. A similar regression analysis of the dependence of the first-order rate, $\Delta G^{\ddagger}k_{chem}$, obtained from biphasic kinetic parameters for fitting time courses for the 28 acylation experiments in Appendix 5.

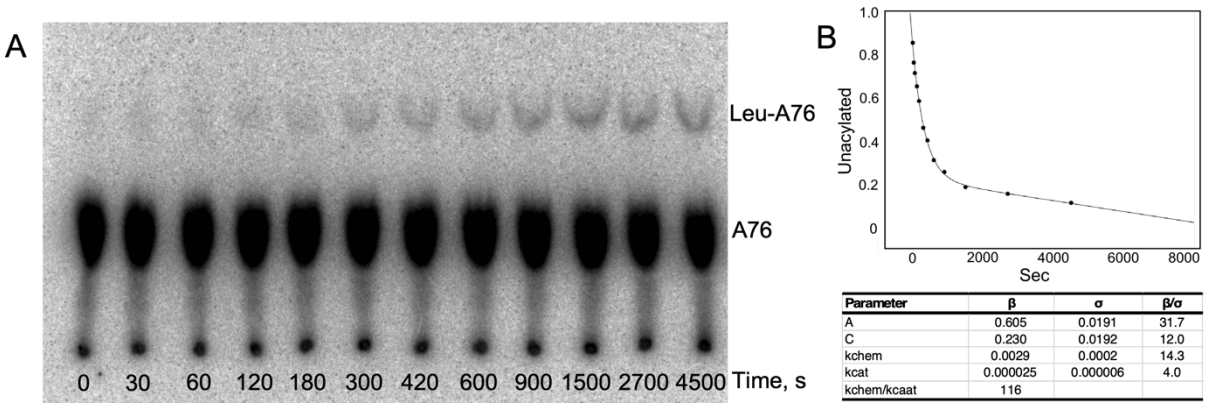


Figure 12. Biphasic fit of a more finely divided time course for LeuAC aminoacylation. The LeuAC used here is the MBP fusion. A. TLC separation of acylated leucyl-2',3'AMP as a function of time. B. Time course in A fitted to a equation (1).

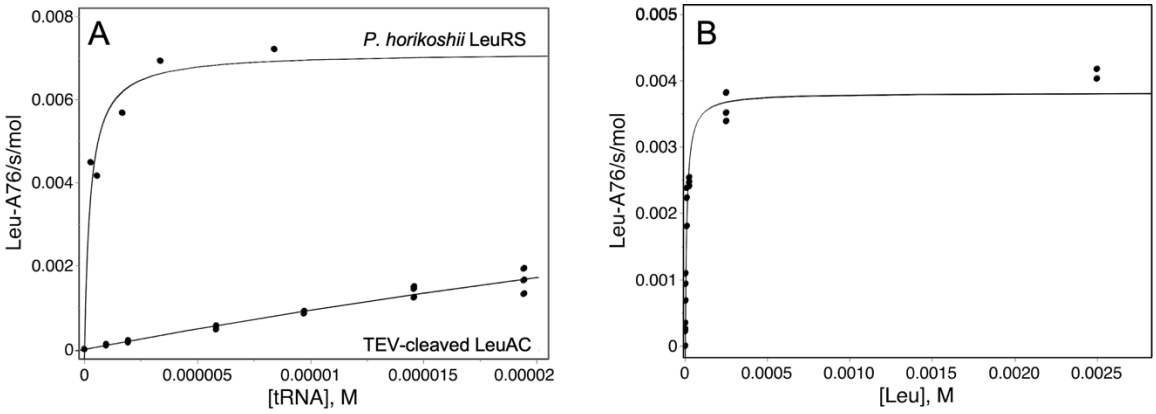
Product release is rate-limiting for both amino acid activation and tRNA^{leu} aminoacylation

Single-turnover analysis of catalytic time courses furnishes a unique estimate for the ratio, (k_{chem}/k_{cat}), of rates for unimolecular conversion of enzyme•substrate to enzyme•product, and product release. In contrast to the [^{32}P] ATP consumption experiments summarized in Figures 7 and 8, which were intentionally active-site titrations and had a 10:3 ratio of ATP to catalysis, the acylation experiments listed in Appendix 5 were done at a wide range of different $tRNA^{Leu}$ and enzyme concentrations. For that reason, distributions of the amplitude parameter, A, and $\Delta G^{\ddagger}(k_{chem}/k_{cat})$ include only experiments in which the [$tRNA^{Leu}$]/[Enzyme] ratios were in the range 1.0 – 5.0 to approximate single-turnover conditions.

Appendices 4 and 5 identify the differences between, and significant contributions to, $\Delta G^{\ddagger}(k_{chem}/k_{cat})$ for LeuRS, MBP-LeuAC, and TEV cleaved LeuAC. Parameters for all three catalysts change consistently between amino acid activation and $tRNA^{Leu}$ aminoacylation. LeuRS and TEV cleaved LeuAC both have k_{chem} values ~55-fold greater than k_{cat}, consistent with the appearance of a pre-steady state burst in aminoacylation. As with amino acid activation, the LeuAC fusion protein is the weakest catalyst (β_{Urz} = +3.9 kcal/mole), but TEV cleavage more than compensates for this (β_{TEV} = –4.7 kcal/mole).

Michaelis-Menten kinetic parameters for LeuAC distinguish its activity from that of LeuRS

Figure 13 shows new experimental Michaelis-Menten data necessary to compare the steady-state dependence of *P. horikoshii* LeuRS and TEV cleaved LeuAC on $tRNA^{Leu}$ and leucine. Despite the obstacles noted elsewhere to steady-state kinetic analysis, these data help to confirm the distinct behavior of the LeuAC catalysis. Saturation of LeuRS and LeuAC by $tRNA^{Leu}$, which were performed using the same tRNA sample, differ by an order of magnitude in K_M. Saturation of LeuAC could not be achieved, owing to the high K_M and the low total acylatability (0.28) of the tRNA substrate. Curvature in the Michaelis-Menten plot can nonetheless be demonstrated by the systematic pattern of residuals when the Michaelis-Menten model is compared to a linear fit. K_M values for leucine-dependence, though similar (LeuAC 9.5±1.3E-6 vs LeuRS 5.9±0.8E-6), also differ by a statistically significant amount (2.8σ).



P. horikoshii LeuRS $tRNA^{Leu}$

Parameter	β	StdErr	β/σ
kcat	0.0071	0.0005	14.5
KM	2.55E-07	9.00E-08	2.8
kcat/KM	2.80E+04		

TEV-cleaved LeuAC $tRNA^{Leu}$

Parameter	β	StdErr	β/σ
kcat	1.00E-02	0.0077	1.30
KM	1.01E-04	9.03E-05	1.12
kcat/KM	99		

TEV-cleaved LeuAC leucine

Parameter	β	StdErr	β/σ
kcat	0.0038	1.20E-04	31.7
KM	9.45E-06	1.26E-06	7.5
kcat/KM	404		

Figure 13. Steady-state kinetic analyses of *P.horikoshii* LeuRS and TEV-Cleaved LeuAC. A. $tRNA^{Leu}$-dependence. Results for LeuRS and LeuAC are plotted on the same graph to emphasize the qualitative difference in K_M. B. Leucine-dependence. Fitted kinetic parameters are given in the tables, together with statistical significance. Measurements for LeuAC were done in triplicate.

Discussion

The notion that far simpler aaRS may have driven early stages in the evolution of translation (15-19) led to enhanced understanding of how the Class distinction results from amino acid physical chemistry (7,8,39), how the operational code in the tRNA acceptor stem dictates aaRS recognition (4-8), and how aaRS secondary structural duality dictates amino acid side chain size discrimination (40). Those byproducts assumed, but are independent of, the authenticity of the reported catalytic activities of the TrpRS and HisRS urzymes and protozymes. Nonetheless, as urzyme catalytic activities underpin the validation of the Rodin-Ohno hypothesis, it is essential to validate and extend them.

Experimental aaRS urzymology presents fundamental difficulties: (i) urzyme catalysis is reduced by 4-5 orders of magnitude relative to that of full-length aaRS, (ii) solubilization of urzymes requires that they be tagged with maltose-binding protein, and (iii) failure to fully define the requirements for activity contributes to poor reproducibility. Such difficulties are characteristic of initial explorations of any previously uninhabited experimental landscape. Metrics compiled from highly replicated enzymatic assays analyzed by approximating single-turnover conditions allowed us to circumvent these difficulties. These metrics, use of [α^{32}P] ATP to monitor all three nucleotides during amino acid activation, and steady-state kinetics of substrate concentration dependence, reveal fundamental differences between the enzymology of the full-length LeuRS and LeuAC. Significant differences between PP_i exchange and $tRNA^{Leu}$ aminoacylation measured for the MBP-LeuAC fusion protein and its TEV cleaved products strengthen the authenticity of the LeuAC-catalyzed reactions.

The leucyl-tRNA synthetase urzyme, LeuAC, appears to be an authentic catalyst of amino acid activation and tRNA acylation

We inferred the authenticity of catalysis by previously described urzymes from the sensitivity of catalysis to genetic manipulation, significantly different K_M values from Michaelis-Menten kinetics experiments, and active-site titration assays using [γ^{32}P] ATP, as described by Fersht (36). The LeuAC urzyme appears to satisfy each of these criteria as an authentic catalyst of both steps in the synthesis of leu-tRNA^{Leu}. Experimental replication has afforded more detailed statistical significance of the resulting evidence than we reported previously, thereby strengthening previous conclusions.

Neither the amplitude values, A, for [^{32}P] ATP consumption nor those for tRNA aminoacylation can be converted into active-site titers, for different reasons. The n-values derived from loss of [^{32}P] ATP are corrupted by the conversion of ATP into ADP (Figures 7B, C). Burst sizes from single-turnover acylation assays are uncertain because the high concentration of inactive tRNA^{Leu} means that an unknown fraction could be inhibitory. Notwithstanding, it is very unlikely that a contaminant present at 0.1 % could produce such large amplitudes.

ATP consumption (Figure 8), by which we previously estimated the active-site titer, pyrophosphate exchange (Figure 10), and tRNA^{Leu} aminoacylation by LeuAC (Figure 11) are all significantly enhanced by TEV cleavage. The enhancement in ATP-dependent activation is about two-fold, compared to an uncertainty of ~0.06, leading to highly significant P-values (1E-10) under the null hypothesis. Aminoacylation of tRNA^{Leu} (Figure 11B, 11C) is similarly increased by TEV cleavage; cleaved LeuAC rates are larger by an order of magnitude, although the

statistical significance (P = 0.003) is lower. Such enhancements cannot be expected from treating contaminating cellular enzymes with TEV protease.

The reciprocal influence of LeuAC and [^{14}C] leucine on each other's elution (Figure 9B) is consistent with a substantive binding stoichiometry (i.e. 0.57 of LeuAC undergoes a mobility shift). However, the mutual displacement of LeuAC and [^{14}C] leucine can support only approximations to leu-5'AMP affinity, because detection required using concentrations substantially higher than the expected dissociation constant. Qualitatively, the ligand off-rate must be slow enough to account for displacement of [^{14}C] leucine equivalent to 85 % of the LeuAC concentration by about half the difference between its unperturbed elution and the protein peak (Figure 9B), suggesting a half-life of up to 10 minutes for the LeuAC•leu-5'AMP complex. For any reasonable on-rate, that half-life sets an upper limit on K_D (i.e., <1E-7 M; see dashed line in Appendix 2) that is two orders of magnitude tighter than the measured K_M value in Figure 9B.

Reactions catalyzed by LeuAC entail near quantitative production of ADP

Results in Figure 7 introduce novel mechanistic questions that we can only begin to address here. Early work by Zamecnik's laboratory documented the catalytic production of noncanonical adenine nucleotides by several aaRS including Class I ValRS, ArgRS and Class II AspRS and PheRS (35). Full-length LeuRS from both *E. coli* and *P. horikoshii* produce minor amounts of ADP along a time-course; in the case of LeuAC, it closely mirrors the time-course for loss of ATP (Figure 7C). In this work, however, the source of the labeled ADP product is [α^{32}P] ATP, rather than either [^{3}H] or [^{32}P] labeled AMP and/or ADP. ADP production thus seems most likely to result from phosphoryl transfer from ATP to enzyme bound aminoacyl 5'-AMP.

Zamecnik's group suggested phosphorylation of the aminoacyl-5'AMP as a major pathway accounting for the appearance of ADP in their experiments, which they interpreted in terms of forming an enzyme-bound thio-acyl intermediate. None of the structural and/or mechanistic studies of aaRS in the intervening years have implicated transient formation of aminoacyl-thioenzyme intermediates in aminoacylation. We therefore propose that the cycle of reactions in Figure 14 reconciles the unexpected data in Figure 7 with the substantial evidence adduced here for LeuAC catalysis of the canonical reactions associated with full-length aaRS.

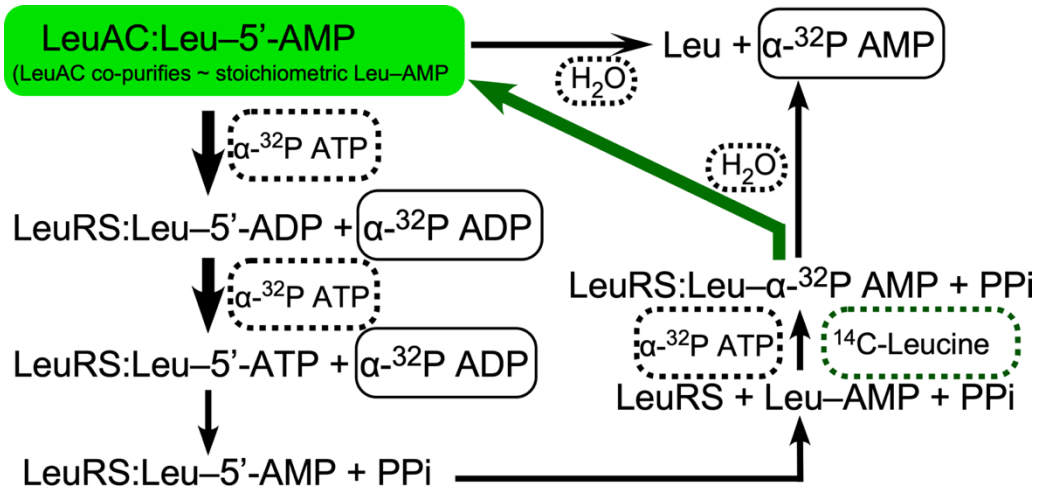


Figure 14. Kinetic scheme accounting for formation of ADP. The non-canonical sur-stoichiometric ADP (bold black arrows) and the incorporation of [^{14}C] leucine into bound leucyl-5'AMP (bold green arrows, green background). Reactants—ATP, leucine, and water—are enclosed by dashed rectangles; products—ADP, AMP, and leucyl-5'AMP—by solid rectangles.

Given that both LeuRS and LeuAC retain substantial amounts of leucyl-5'AMP, that ligand is likely the initial state of both catalysts. We propose that the adenylate is more exposed in LeuAC than it is in LeuRS, and that the specificity for the tRNA^{Leu} 3'-terminal adenosine binding pocket in LeuAC is relaxed sufficiently to accommodate ATP in a manner conducive to phosphorylation of the bound adenylate (Figure 14 thick black arrows), consistent with the near

simultaneous, stoichiometric consumption of ATP with production of ADP (Figure 7C). The amplitude of AMP formation (~5 % of total LeuAC; Figure 8B) suggests that the off-rate for the leucyl-5'AMP is sufficient to enable it to dissociate, leading rapidly to hydrolysis and rebinding of [^{14}C] leucine under the saturating concentrations of ATP and leucine, and to incorporation of the label as [α^{32}P] AMP into the adenylate (Figure 14 thick green arrows), in keeping with our interpretation of Figure 9B.

The mechanistic complications represented in Figure 14 will require further work, especially considering the two-way interactions in Figure 10B, C. Inhibition by KF of [^{32}P] ATP synthesis from [^{32}P] PP_i by LeuAC, but not by LeuRS (Figure 10B) suggests that phosphoryl-transfer from ATP to the bound adenylate may participate in the overall catalytic cycle. The sign of the LeuAC-leucine interaction term (Figure 10C) implies that the binding of leucine to LeuAC stimulates PP_i exchange less effectively in LeuAC than in the full-length enzyme. Among the clues that must be followed up are (i) that AMP is generated ~20 times faster by LeuRS than by LeuAC in single-turnover experiments (Appendix 3A,B,D), (ii) the presence of a third phase with increased apparent turnover in the time course for AMP production by LeuAC (Appendix 3C), and the observation that high ADP concentrations inhibit catalysis by LeuAC (Zhijie Li unpublished data).

Realizing the potential utility of aaRS urzymes will require more thoughtful redesign

The dominant factor limiting more detailed studies of aaRS urzymes, especially structural studies (29), is their limited solubility. Much of our early effort to address this problem grew out of a curious N-terminal packet of charges in LeuAC (Figure 3). TrpRS chimeras containing this N-terminal extension were indeed substantially more soluble, but also inactive. Data reported

here for the LeuAC urzyme, together with Appendix 1, suggest that soluble, active urzymes must be engineered more carefully, perhaps using newer and more effective Rosetta algorithms (41,42).

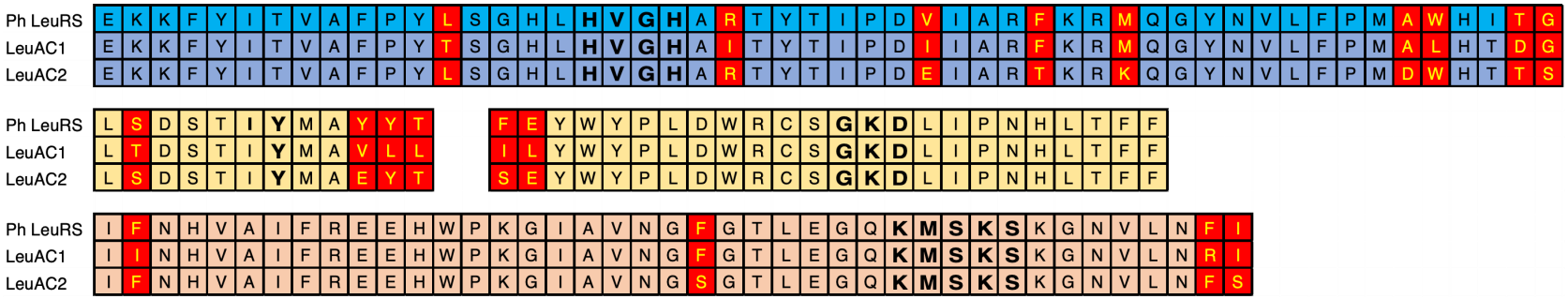
APPENDIX 1: GENE AND PROTEIN SEQUENCES OF LEUAC CONSTRUCTS

LeuAC_1:

```
GAG AAA AAA TTC TAC ATC ACC GTG GCG TTC CCG TAC ACG AGC GGC CAT CTG CAC GTT GGT CAC GCG ATT ACC TAT ACG ATC
CCG GAT ATT ATC GCC CGT TTT AAA CGC ATG CAG GGC TAC AAT GTG CTG TTC CCG ATG GCC CTG CAT ACC GAT GGT CTG ACC
GAT AGT ACG ATT TAT ATG GCA GTT CTG CTG ATC CTG TAT TGG TAC CCG CTG GAT TGG CGT TGC AGC GGC AAA GAT CTG ATT
CCG AAC CAT CTG ACC TTT TTC ATC ATC AAC CAC GTG GCA ATC TTC CGC GAA GAA CAT TGG CCG AAA GGT ATC GCG GTT AAC
GGC TTC GGT ACG CTG GAA GGC CAG AAA ATG AGC AAA TCT AAG GGT AAC GTG CTG AAT CGT ATC
```

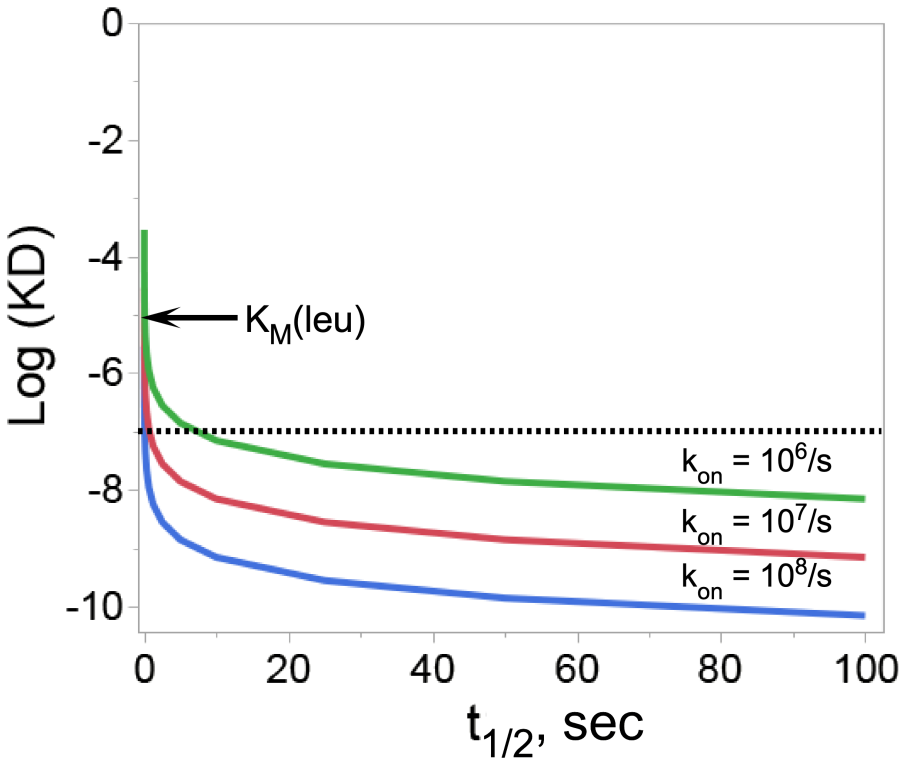
LeuAC_2:

```
GAA AAG AAA TTT TAT ATC ACC GTG GCC TTT CCG TAT CTG AGT GGC CAT CTG CAT GTT GGT CAT GCC CGC ACC TAT ACC ATT
CCG GAT GAA ATT GCA CGC ACC AAA CGT AAA CAG GGC TAT AAT GTT CTG TTT CCG ATG GAT TGG CAT ACC ACC AGC CTG AGC
GAT AGC ACC ATC TAT ATG GCA GAA TAT ACC AGT GAA TAT TGG TAT CCG CTG GAT TGG CGC TGC AGC GGC AAA GAT CTG ATT
CCG AAT CAT CTG ACC AAA TTC ATT TTT AAT CAC GTG GCA ATT TTC CGT GAA GAA CAT TGG CCG AAA GGT ATT GCC GTT AAT
GGC AGT GGT ACA CTG GAA GGC CAG AAA ATG AGT AAA AGC AAA GGT AAT GTT CTG AAT TTC AGC
```

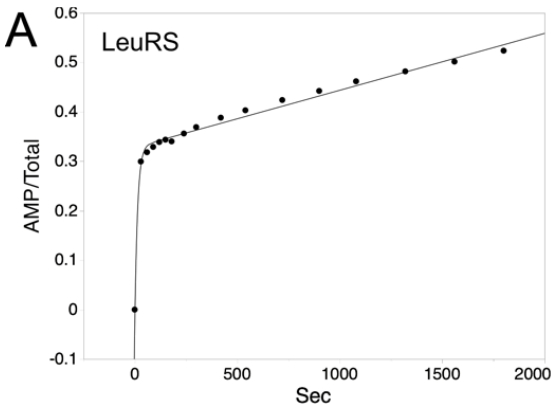
Sequence differences between LeuAC constructs and homologous sequence from *P. horikoshii* LeuRS. The sequence is broken into three fragments analogous to the three described by Pham, et al., 2007. Connecting peptide 1 (CP1) connects the C-terminus of the blue fragment to the N-terminus of the Amber fragment. Connecting Peptide 2 (CP2) is indicated by the gap in the Amber fragment. Differences between the three sequences are highlighted in red boxes. Active-site residues are highlighted in bold and larger font size. Entries for LeuAC2 enhance the solubility of LeuAC2 but have no detectable effect on the active-site titration time courses.

APPENDIX 2: SIMULATED K_D OF LEUAC LEUCYL-5'AMP

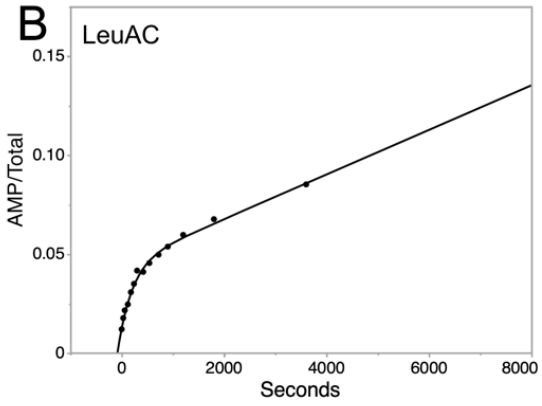


Simulation of dissociation constants consistent with the displacement of bound [^{14}C] leucine in Figure 9B. Assuming that the dissociation constant, $K_D = k_{off}/k_{on}$, $\log(K_D)$ can be plotted against the half-life of the complex, $t_{1/2} = 0.693*k_{off}$, for reasonable values of k_{on}. The observed displacement (Figure 9B) represents approximately 10 minutes or 40% of the total transit time in the G-15 support. The abscissa covers an ~100-fold range ending at 100 seconds, or 1/6th of the actual displacement. Simulated curves are shown for three possible on-rates. Essentially all possible values for K_D are less than 10^{-7} M, which is two orders of magnitude tighter than the K_M for leucine in Figure 13B.

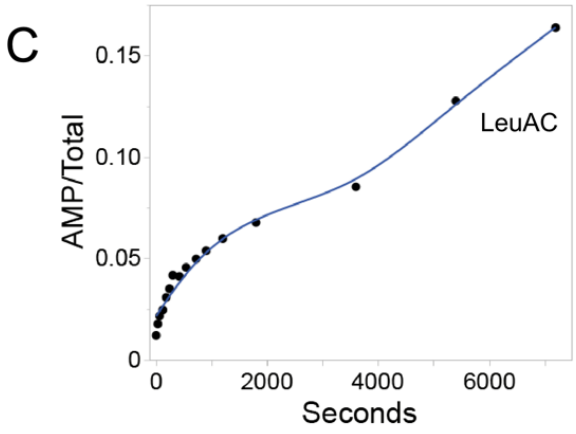
APPENDIX 3: LEURS AND LEUAC AMP FORMATION

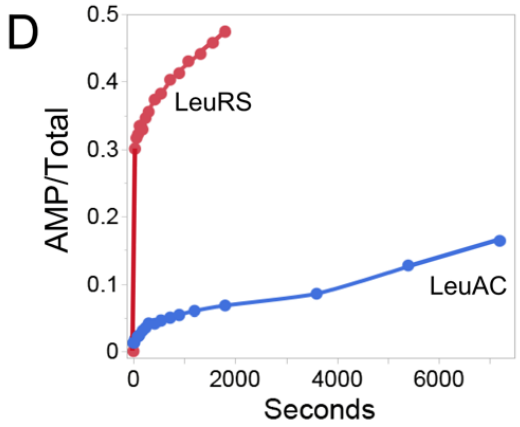


Parameter	β	StdErr	β/σ
C	0.32	0.00375	86.5
A	-0.32	0.01003	-32.3
kchem	8.1E-02	0.012	7.0
kcat	9.0E-05	4.40E-06	20.4
kchem/kcat	902		



Parameter	β	StdErr	β/σ
C	0.05	0.00215	21.1
A	-0.03	0.0022	-14.6
kchem	3.61E-03	0.00056	6.5
kcat	1.13E-5	9.51E-07	1.19E+01
kchem/kcat	319		





Time courses for AMP formation by LeuRS and LeuAC. A. Active-site titration plot for LeuRS. B. Active-site titration plot for LeuAC, using only timepoints to 4000 seconds. C. Extended time course for LeuAC, showing an increased linear turnover rate from 4000 to 7200 seconds. D. LeuRS and LeuAC time courses compared on the same coordinate system.

APPENDIX 4: BIPHASIC (SINGLE-TURNOVER) FITS OF [^{32}P] ATP LOSS

Sample	n	Date	C	A	kchem	kcat	kchem/ kcat	ΔGkchem	ΔGkcat	ΔGkchem/ kcat	URZ	TEV	α labeled	LSA	AMPcPP	*P. horik*
Ph LeuRS	1.2	6/2/21	0.68	0.32	0.08085	8.97E-05	902	1.49	5.52	-4.03	0	0	0	0	0	1
Ph LeuRS	1.3	5/25/21	0.63	0.37	0.06870	1.59E-04	431	1.59	5.18	-3.59	0	0	0	0	0	1
Ec LeuRS	1.0	12/14/20	0.13	0.29	0.00178	7.06E-06	252	3.75	7.02	-3.27	0	0	0	0	0	0
Ec LeuRS	1.6	12/14/20	0.13	0.44	0.00052	4.30E-06	120	4.48	7.32	-2.83	0	0	1	0	1	0
Ec LeuRS	1.4	12/14/20	0.29	0.39	0.00082	4.42E-06	185	4.21	7.30	-3.09	0	0	1	1	0	0
Ec LeuRS	0.9	12/14/20	0.12	0.24	0.00236	6.05E-06	389	3.58	7.11	-3.53	0	0	1	0	0	0
Ec LeuRS	1.0	12/14/20	0.13	0.29	0.00178	7.06E-06	252	3.75	7.02	-3.27	0	0	1	0	0	0
LeuAC (MBP)	2.4	8/9/18	0.26	0.68	0.00043	2.18E-05	20	4.59	6.35	-1.76	1	0	0	0	0	1
LeuAC (MBP)	2.4	5/25/21	0.26	0.68	0.00043	2.18E-05	20	4.59	6.35	-1.76	1	0	0	0	0	1
LeuAC (MBP)	2.2	12/14/20	0.43	0.62	0.00051	1.19E-05	43	4.49	6.71	-2.22	1	0	1	0	1	1
LeuAC (MBP)	2.5	12/14/20	0.37	0.69	0.00036	4.43E-06	82	4.69	7.30	-2.61	1	0	1	0	0	1
LeuAC (MBP)	2.0	12/4/20	0.42	0.57	0.00033	4.36E-06	75	4.75	7.31	-2.55	1	0	1	0	0	1
LeuAC (MBP)	2.3	12/18/20	0.35	0.65	0.00040	4.27E-06	93	4.64	7.32	-2.68	1	0	1	0	0	1
LeuAC (MBP)	2.5	12/18/20	0.31	0.69	0.00036	*ND*	*ND*	4.69	*ND*	*ND*	1	0	1	0	0	1
LeuAC (MBP)	2.4	12/18/20	0.34	0.67	0.00038	*ND*	*ND*	4.67	*ND*	*ND*	1	0	1	0	0	1
LeuAC (MBP)	2.4	12/18/20	0.33	0.67	0.00037	*ND*	*ND*	4.67	*ND*	*ND*	1	0	1	0	0	1
LeuAC (MBP)	2.3	12/18/20	0.37	0.64	0.00041	*ND*	*ND*	4.62	*ND*	*ND*	1	0	1	0	0	1
LeuAC (MBP)	2.3	12/18/20	0.35	0.65	0.00042	*ND*	*ND*	4.60	*ND*	*ND*	1	0	1	0	0	1
LeuAC (MBP)	2.4	12/18/20	0.34	0.66	0.00037	*ND*	*ND*	4.68	*ND*	*ND*	1	0	1	0	0	1
LeuAC (TEV)	1.9	5/25/21	0.41	0.52	0.00123	2.33E-05	53	3.97	6.31	-2.35	1	1	0	0	0	1
LeuAC (TEV)	2.1	5/4/21	0.27	0.60	0.00326	2.00E-05	161	3.39	6.41	-3.01	1	1	1	0	0	1
LeuAC (TEV)	2.2	5/4/21	0.26	0.62	0.00321	2.07E-05	155	3.40	6.38	-2.99	1	1	1	0	0	1
LeuAC (TEV)	2.2	10/19/17	0.42	0.61	0.00070	1.30E-05	52	4.30	6.66	-2.34	1	1	0	0	0	1

LeuAC (TEV)	2.1	10/19/17	0.43	0.59	0.00066	1.80E-05	37	4.34	6.47	-2.14	1	1	0	0	0	1
LeuAC (TEV)	2.1	5/4/21	0.27	0.58	0.00331	1.97E-05	168	3.38	6.41	-3.03	1	1	0	0	0	1
LeuAC (TEV)	1.9	8/9/18	0.41	0.52	0.00123	2.33E-05	53	3.97	6.31	-2.35	1	1	0	0	0	1

ND denotes kcat values that were determined with $\beta/\sigma < 1.0$

APPENDIX 5: BIPHASIC (SINGLE-TURNOVER) FITS OF AMINOACYLATION EXPERIMENTS

Sample	Date	C	A	kchem	kcat	kchem/ kcat	ΔGkchem	ΔGkcat	ΔGkchem/ kcat	URZ	TEV	[tRNA], M	[enzyme], M	tRNA/ E	14CLeu
Ph LeuRS	9/30/20	0.012	0.90	0.0083	1.77E-05	466	2.84	6.48	-3.64	0	0	5.00E-06	5.00E-07	10.0	0
Ph LeuRS	9/24/20	0.001	0.97	0.0026	3.10E-05	83	3.54	6.15	-2.61	0	0	1.00E-05	5.00E-06	2.0	0
Ph LeuRS	9/24/20	0.291	0.68	0.0036	1.53E-04	23	3.34	5.20	-1.87	0	0	1.00E-05	5.00E-06	2.0	0
Ph LeuRS	10/5/20	0.355	0.32	0.0100	3.80E-04	27	2.72	4.66	-1.95	0	0	5.33E-06	5.00E-06	1.1	0
Ph LeuRS	11/23/20	0.084	0.93	0.0046	3.51E-05	130	3.19	6.07	-2.88	0	0	3.24E-05	5.00E-06	6.5	1
Ph LeuRS	11/17/20	0.694	0.31	0.0181	3.19E-04	57	2.38	4.77	-2.39	0	0	1.60E-05	2.60E-05	0.6	1
Ph LeuRS	10/7/20	0.056	0.94	0.0031	3.96E-05	77	3.43	6.00	-2.58	0	0	5.00E-06	1.25E-05	0.4	0
Ph LeuRS	10/7/20	0.137	0.86	0.0041	7.18E-05	57	3.26	5.65	-2.39	0	0	5.00E-06	1.25E-05	0.4	0
Ph LeuRS	10/7/20	0.092	1.15	0.0051	6.68E-05	77	3.13	5.69	-2.57	0	0	4.30E-05	1.25E-05	3.4	0
Ph LeuRS	10/7/20	0.278	0.49	0.0046	1.45E-04	31	3.19	5.23	-2.04	0	0	4.30E-05	1.25E-05	3.4	0
Ph LeuRS	10/7/20	0.356	0.53	0.0041	1.85E-04	22	3.26	5.09	-1.83	0	0	4.30E-05	1.25E-05	3.4	0
Ph LeuRS	10/28/20	0.096	0.90	0.0538	1.97E-04	274	1.73	5.05	-3.32	0	0	2.60E-05	1.30E-05	2.0	1
Ph LeuRS	10/28/20	0.201	0.80	0.0819	2.01E-04	407	1.48	5.04	-3.56	0	0	2.60E-05	1.30E-05	2.0	0
Ph LeuRS	10/9/20	0.292	0.70	0.0440	2.43E-04	179	1.86	4.93	-3.07	0	0	4.60E-08	1.25E-05	0.0	0
Ph LeuRS	10/9/20	0.384	0.60	0.0240	1.84E-04	128	2.22	5.09	-2.87	0	0	1.67E-07	1.25E-05	0.0	0
Ph LeuRS	10/9/20	0.044	0.91	0.0150	7.00E-05	219	2.48	5.67	-3.19	0	0	2.80E-06	1.25E-05	0.2	0
Ph LeuRS	10/6/20	0.016	1.01	0.0092	5.28E-06	1744	2.78	7.19	-4.42	0	0	5.33E-06	5.00E-06	1.1	0
Ph LeuRS	10/6/20	-0.021	0.90	0.0053	6.32E-06	831	3.11	7.09	-3.98	0	0	5.33E-06	5.00E-06	1.1	0
Ph LeuRS	10/6/20	0.180	0.61	0.0063	9.82E-05	64	3.01	5.46	-2.46	0	0	5.33E-06	5.00E-06	1.1	0
Ph LeuRS	10/6/20	-0.153	0.75	0.0053	6.00E-05	88	3.11	5.75	-2.65	0	0	5.33E-06	5.00E-06	1.1	0
LeuAC (MBP)	9/4/19	0.379	0.55	0.0007	9.00E-05	8	4.32	5.52	-1.20	1	0	6.00E-05	5.00E-06	12.0	0
LeuAC (MBP)	3/12/21	0.191	0.81	0.0010	5.48E-05	18	4.11	5.81	-1.70	1	0	9.13E-05	9.10E-06	10.0	0
LeuAC (MBP)	3/24/19	0.230	0.60	0.0029	2.50E-05	116	3.46	6.27	-2.82	1	0	6.00E-05	5.00E-06	12.0	0

LeuAC (MBP)	4/26/21	0.157	0.85	0.0023	5.91E-05	38	3.61	5.76	-2.16	1	0	8.02E-05	3.60E-05	2.2	0
LeuAC (TEV)	3/12/21	-0.017	1.02	0.0021	5.44E-06	387	3.65	7.18	-3.53	1	1	9.13E-05	5.40E-06	16.9	0
LeuAC (TEV)	2/25/19	0.414	0.34	0.0039	3.00E-04	14	3.28	4.84	-1.56	1	1	1.50E-05	5.00E-06	3.0	0
LeuAC (TEV)	3/29/19	0.526	0.47	0.0270	2.48E-04	108	2.15	4.92	-2.77	1	1	5.00E-06	5.00E-06	1.0	0
LeuAC (TEV)	3/22/19	0.258	0.74	0.0491	8.47E-04	58	1.79	4.19	-2.40	1	1	6.00E-05	5.00E-06	12.0	0

REFERENCES

1. Gilbert, W. (1986) Origin of Life: The RNA world. *Nature*, **319**, 618.

2. Carter, C.W., Jr. (2015) What RNA World? Why a Peptide/RNA Partnership Merits Renewed Experimental Attention. *Life*, **5**, 294-320.

3. Eigen, M. (1971) Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften,* **58,** 465–523.

4. Carter, C.W., Jr and Wills, P.R. (2019) Class I and II aminoacyl-tRNA synthetase tRNA groove discrimination created the first synthetase•tRNA cognate pairs and was therefore essential to the origin of genetic coding. *IUBMB Life*, **71**, 1088–1098.

5. Carter, C.W., Jr and Wills, P.R. (2018) Hierarchical groove discrimination by Class I and II aminoacyl-tRNA synthetases reveals a palimpsest of the operational RNA code in the tRNA acceptor-stem bases. *Nucleic Acids Res.*, **46**, 9667–9683.

6. Carter, C.W., Jr. and Wills, P.R. (2021) The Roots of Genetic Coding in Aminoacyl-tRNA Synthetase Duality. *Ann. Rev. Biochem.*, **90**, 349-373.

7. Carter, C.W., Jr. and Wolfenden, R. (2016) Acceptor-stem and anticodon bases embed amino acid chemistry into tRNA. *RNA Biol.*, **13**, 145–151.

8. Carter, C.W., Jr. and Wolfenden, R. (2015) tRNA Acceptor-Stem and Anticodon Bases Form Independent Codes Related to Protein Folding. *Proc. Nat. Acad. Sci. USA*, **112**, 7489-7494.

9. Francklyn, C.S., First, E.A., Perona, J.J., Hou, Y. (2008) Methods for kinetic and thermodynamic analysis of aminoacyl-tRNA synthetases. *Methods,* 44, Issue 2, 100-118.

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

11. Nagel, G. M., & Doolittle, R. F. (1991) Evolution and relatedness in two aminoacyl-tRNA synthetase families. *Proc. Nat. Acad. Sci.,* **88**, 8121-8125.

12. Rodin, S., Ohno, S., & Rodin, A. (1993) On concerted origin of transfer RNAs with complementary anticodons. *Origins Life Evol. Biosphere,* 23, 393-418.

13. Wong, J. T. F. (1975) A co-evolution theory of the genetic code. *Proc. Nat. Acad. Sci. USA,* **72.** 1909.

14. Rodin, S.N. and Ohno, S. (1995) Two types of aminoacyl-trna synthetases could be originally encoded by complementary strands of the same nucleic ACID. *Origins Life Evol. Biosphere,* **25,** 565–589.

15. Pham, Y., Li, L., Kim, A., Weinreb, V., Butterfoss, G., Kuhlman, B. and Carter, C.W., Jr. (2007) A Minimal TrpRS Catalytic Domain Supports Sense/Antisense Ancestry of Class I and II Aminoacyl-tRNA Synthetases. *Mol. Cell*, **25**, 851-862.

16. Pham, Y., Kuhlman, B., Butterfoss, G.L., Hu, H., Weinreb, V. and Carter, C.W., Jr. (2010) Tryptophanyl-tRNA synthetase Urzyme: a model to recapitulate molecular evolution and investigate intramolecular complementation. *J. Biol. Chem.*, **285**, 38590-38601.

17. Li, L., Weinreb, V., Francklyn, C. and Carter, C.W., Jr. (2011) Histidyl-tRNA Synthetase Urzymes: Class I and II Aminoacyl-tRNA Synthetase Urzymes have Comparable Catalytic Activities for Cognate Amino Acid Activation. *J. Biol. Chem.*, **286**, 10387-10395.

18. Li, L., Francklyn, C. and Carter, C.W., Jr. (2013) Aminoacylating Urzymes Challenge the RNA World Hypothesis. *J. Biol. Chem.*, **288**, 26856-26863.

19. Martinez, L., Jimenez-Rodriguez, M., Gonzalez-Rivera, K., Williams, T., Li, L., Weinreb, V., Chandrasekaran, S.N., Collier, M., Ambroggio, X., Kuhlman, B. *et al.* (2015) Functional Class I and II Amino Acid Activating Enzymes Can Be Coded by Opposite Strands of the Same Gene. *J. Biol. Chem.*, **290**, 19710–19725.

20. Carter, C.W., Jr. (2014) Urzymology: Experimental Access to a Key Transition in the Appearance of Enzymes. *J. Biol. Chem.*, **289**, 30213–30220.

21. Khlebnikov, A., Datsenko, K.A., Skaug, T., Wanner, B.L., Keasling, J.D. (2001) Homogeneous expression of the P(BAD) promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiol.*, **147**, 3241-3247.

22. Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G., Court, D.L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc. Nat. Acad. Sci. USA*, **97**, 5978-83.

23. Thomason, L.C., Costantino, N., Court, D.L. (2007) E. coli genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* Chapter 1.17.

24. Tiruvadi Krishnan, S., Moolman, M. C., van Laar, T., Meyer, A. S., Dekker, N. H. (2015) Essential validation methods for E. coli strains created by chromosome engineering. *J.ournal of biol. engineering*, ***9*,** 1-14.

25. Badran, A.H. and Liu, D.R. (2015) Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.*, **6**, 1-10.

26. Trevors, J. (1986) Plasmid curing in bacteria. *FEMS Microbiol. Letters,* **32**, 149-157.

27. Clark, A.J. (1973) Recombination deficient mutants of E. coli and other bacteria. *Annu. Rev. Genet.*, **7**, 67–86.

28. Onodera, K., Suganuma, N., Takano, H., Sugita, Y., Shoji, T., Minobe, A., Yamaki, N., Otsuka, R., Mutsuro-Aoki, H., Umehara, T. *et al.* (2021) Amino acid activation analysis of primitive aminoacyl-tRNA synthetases encoded by both strands of a single gene using the malachite green assay. *BioSystems*, **208**, 104481.

29. Sapienza, P.J., Li, L., Williams, T., Lee, A.L. and Carter, C.W., Jr. (2016) An Ancestral Tryptophanyl-tRNA Synthetase Precursor Achieves High Catalytic Rate Enhancement without Ordered Ground-State Tertiary Structures. *ACS Chem Biol.*, **11**, 1661−1668.

30. Boniecki, M.T., Vu, M.T., Betha, A.K. and Martinis, S.A. (2008) CP1-dependent partitioning of pretransfer and posttransfer editing in leucyl-tRNA synthetase. *Proc. Nat. Acad. Sci. USA,* **105**, 19223–19228.

31. Vu, M.T. and Martinis, S.A. (2007) A Unique Insert of Leucyl-tRNA Synthetase Is Required for Aminoacylation and Not Amino Acid Editing. *Biochem.*, **46**, 5170-5176.

32. Lincecum Jr., T.L., Tukalo*, M., Yaremchuk*, A., Mursinna, R.S., Williams, A.M., Sproat, B.S., Van Den Eynde, W., Link, A., Calenbergh, S. V., Grotli, M., Martinis*, S.A. and Cusack*, S. (2003) Structural and Mechanistic Basis of Pre- and Post-Transfer Editing by Leucyl-tRNA Synthetase. *Mol Cell,* **11**, 951-963.

33. Prætorius-Ibba, M., Rogers, R., Samson, T.E., Kelman, Z. and Ibba, M. (2005) Association between Archaeal Prolyl- and Leucyl-tRNA Synthetases Enhances tRNAPro Aminoacylation. *J. Biol. Chem.*, **280**, 26099–26104.

34. Cvetesic, N., Perona, J.J. and Gruic-Sovulj, I. (2012) Kinetic Partitioning between Synthetic and Editing Pathways in Class I Aminoacyl-tRNA Synthetases Occurs at Both Pre-transfer and Post-transfer Hydrolytic Steps. *J. Biol. Chem.*, **287**, 25381–25394.

35. Rapaport, E., Remy, P., Kleinkauf, H., Vater, J. and Zamecnik, P.C. (1987) Aminoacyl-tRNA synthetases catalyze AMP -* ADP ATP exchange reactions, indicating labile covalent enzyme-amino acid intermediates (protein synthesis/aminoacyl adenylates/aminoacylation. *Proc. Nati. Acad. Sci. USA*, **84**, 7891-7895.

36. Fersht, A.R., Ashford, J.S., Bruton, C.J., Jakes, R., Koch, G.L.E. and Hartley, B.S. (1975) Active-site Titration and Aminoacyl Adenylate Binding Stoichiometry of Amionacyl-tRNA Synthetases. *Biochem.*, **14**, 1-4.

37. Berger, G. and Girault, G. (2003) Macromolecule–ligand binding studied by the Hummel and Dreyer method: current state of the methodology. *J. Chromat. B*, **797**, 51–61.

38. Phillips, M.L. and White, R.L. (1997) Dependence of Chromatogram Peak Areas Obtained by Curve-Fitting on the Choice of Peak Shape Function. *J. Chromat. Sci.* **35,** 75-81.

39. Wolfenden, R., Lewis, C.A., Yuan, Y. and Carter, C.W., Jr. (2015) Temperature dependence of amino acid hydrophobicities. *Proc. Nat. Acad. Sci. USA*, **112**, 7484-7488.

40. Carter, C.W., Jr. and Wills, P.R. (2019) Experimental Solutions to Problems Defining the Origin of Codon-Directed Protein Synthesis. *BioSystems*, **183**, 103979.

41. Jacak, R.J., Leaver-Fay, A.T. and Kuhlman, B. (2012) Computational protein design with explicit consideration of surface hydrophobic patches. *Proteins*, **80**, 825–838.

42. Kuhlman, B. and Bradley, P. (2019) Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.*, **20**, 681–697.