

# Domain acquisition by class I aminoacyl-tRNA synthetase urzymes coordinated the catalytic functions of HVGH and KMSKS motifs

Guo Qing Tang<sup>1</sup>, Jessica J.H. Elder<sup>1</sup>, Jordan Douglas<sup>1,2</sup> and Charles W. Carter, Jr<sup>1,\*</sup>

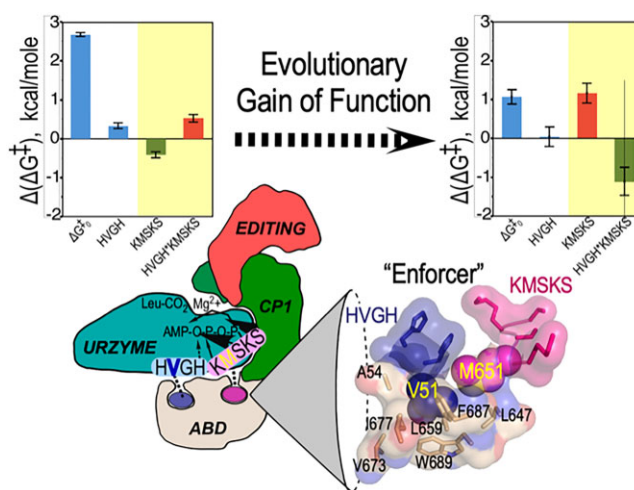
<sup>1</sup>Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599-7260, USA and  
<sup>2</sup>Department of Physics, The University of Auckland, New Zealand

Received April 11, 2023; Revised June 23, 2023; Editorial Decision June 23, 2023; Accepted July 11, 2023

## ABSTRACT

Leucyl-tRNA synthetase (LeuRS) is a Class I aminoacyl-tRNA synthetase (aaRS) that synthesizes leucyl-tRNA<sup>leu</sup> for codon-directed protein synthesis. Two signature sequences, HxGH and KMSKS help stabilize transition-states for amino acid activation and tRNA aminoacylation by all Class I aaRS. Separate alanine mutants of each signature, together with the double mutant, behave in opposite ways in *Pyrococcus horikoshii* LeuRS and the 129-residue urzyme ancestral model generated from it (LeuAC). Free energy coupling terms,  $\Delta(\Delta G^\ddagger)$ , for both reactions are large and favourable for LeuRS, but unfavourable for LeuAC. Single turnover assays with <sup>32</sup>P $\alpha$ -ATP show correspondingly different internal products. These results implicate domain motion in catalysis by full-length LeuRS. The distributed thermodynamic cycle of mutational changes authenticates LeuAC urzyme catalysis far more convincingly than do single point mutations. Most importantly, the evolutionary gain of function induced by acquiring the anticodon-binding (ABD) and multiple insertion modules in the catalytic domain appears to be to coordinate the catalytic function of the HxGH and KMSKS signature sequences. The implication that backbone elements of secondary structures achieve a major portion of the overall transition-state stabilization by LeuAC is also consistent with coevolution of the genetic code and metabolic pathways necessary to produce histidine and lysine sidechains.

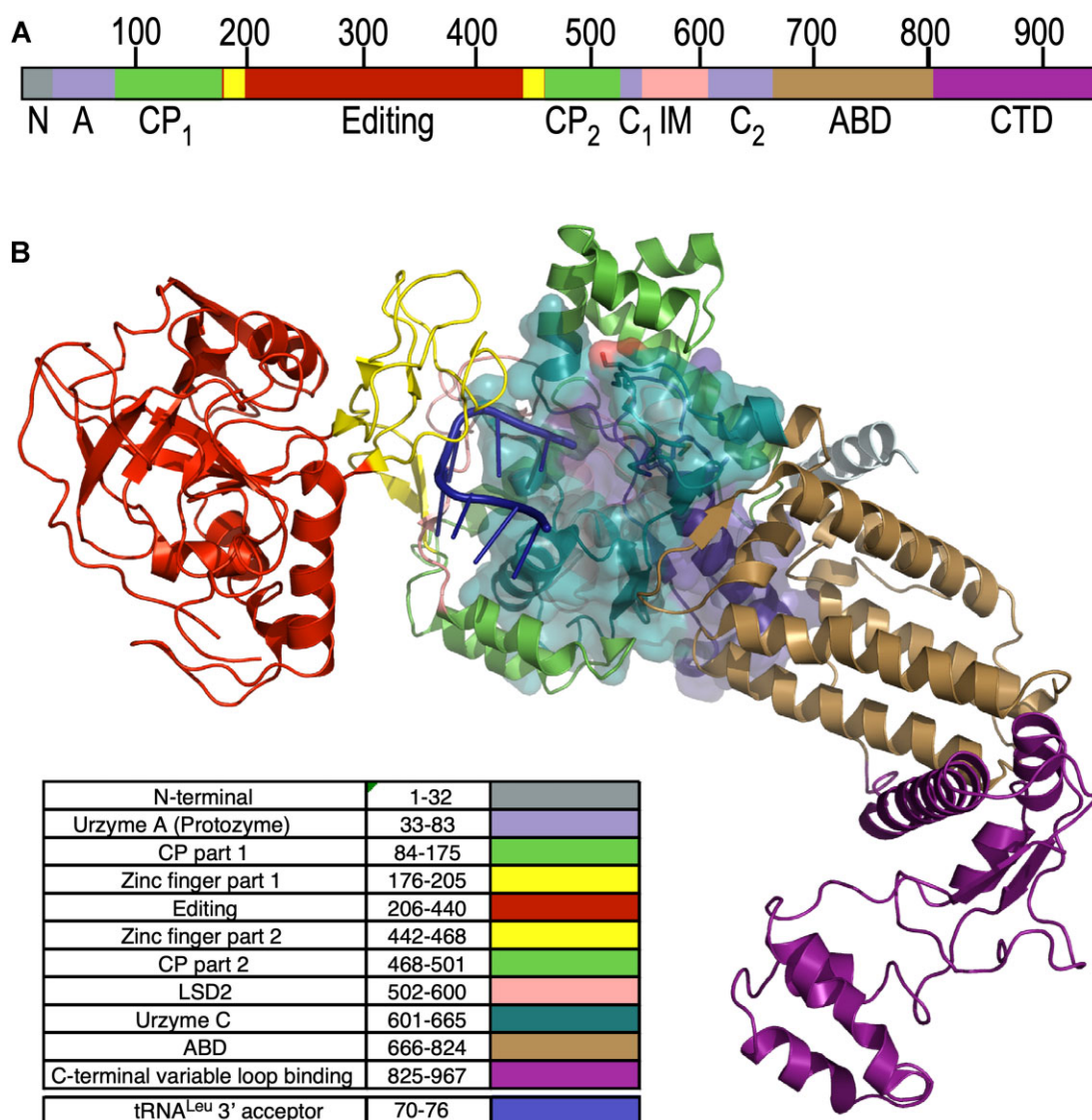
## GRAPHICAL ABSTRACT



## INTRODUCTION

Class I aminoacyl-tRNA synthetases (aaRS) are one of two enzyme superfamilies that, together with their cognate tRNAs, translate the genetic code (1). They perform that function by activating cognate amino acids at the expense of the two high-energy bonds in ATP and transferring the activated aminoacyl group from the 5' position of AMP to the tRNA 3'-terminal ribose. Eleven contemporary Class I aaRS families are responsible for the hydrophobic amino acids (isoleucine, valine, leucine, methionine), together with the larger homologs of the acidic (glutamate), amide (glutamine), and aromatic (tryptophan, tyrosine) amino acids, as well as arginine and cysteine. Some archaea and bacteria also possess a Class I lysyl-tRNA synthetase (2), which in most organisms is a Class II enzyme. All Class I aaRS share the same three domains (Figure 1): a Rossmann dinucleotide-binding catalytic domain, variable-length insertions between the two crossover connections of

\*To whom correspondence should be addressed. Tel: +1 919 259 2558; Fax: +1 919 966 2852; Email: carter@med.unc.edu  
Present address: Jessica J. H. Elder, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA.



**Figure 1.** Relation of LeuAC urzyme to full-length *P. horikoshii* LeuRS. Colors are consistent between A and B. (A) Linear schematic of LeuRS domains. N, N-terminal helical domain; A, Protozyme (ATP-binding site); CP, connecting peptide, a nested insertion containing the editing domain; Editing, Editing domain; C<sub>1</sub>, second fragment of the urzyme (amino acid and tRNA binding). LSD<sub>2</sub>, Leucine specific domain 2 (11), an insertion module before second crossover; C<sub>2</sub>, C-terminal fragment of the urzyme (pyrophosphate binding); ABD, anticodon-binding domain; CTD, C-terminal domain that binds to the tRNA<sup>Leu</sup> variable loop. (B) Three-dimensional cartoon based on PDB ID 1WC2 showing the LeuAC urzyme (surface embedded into the full-length enzyme and the arrangement of modules acquired during the specialization of LeuRS from other Class I aaRS. A 76 of the tRNA<sup>Leu</sup> acceptor stem inserts into the active sites of both LeuRS and LeuAC, bisecting the protozyme (blue) and urzyme part C (C<sub>2</sub>), the second half of the urzyme. Insertions into a loop connecting the protozyme and C2 are all nested, one within the next, forming the complete CP. In contrast, C-terminal additions are serial.

the Rossmann fold (CP) (3,4) including an embedded editing domain in larger Class I enzymes, and an idiosyncratic anticodon-binding domain, ABD. The active site within the Rossmann fold is highly conserved across the superfamily (5). Moreover, that domain has been re-constructed without any of the other domains, and it retains all three functions—amino acid activation, retention of the activated amino acid, and transfer to cognate tRNA—necessary to implement a rudimentary form of coding. For these reasons, we have argued that such constructs, called ‘urzymes’ (6), are excellent experimental models for studying the emergence and early evolutionary history of ancestral aaRS (7–10) and genetic coding generally.

Here, we utilize the leucyl-tRNA synthetase from *Pyrococcus horikoshii*, LeuRS (11) and its urzyme, LeuAC (12), to examine the evolutionary gain of function associated with acquiring the CP and ABD domains. We note that as *P. horikoshii* expresses the archaeal-like LeuRS, which is monomeric, LeuAC is also very likely monomeric. The location of its post-transfer editing domain within the extended connecting peptide (CP) insertion differs from that in eubacterial LeuRS (13). Construction of the LeuAC urzyme entailed removing the entire extended CP domain, which includes many acquired insertion modules (Douglas, J., Bouckaert, R., Carter, C.W.J. and Wills, P. (2023) Enzymic recognition of amino acids drove the

evolution of primordial genetic codes. *Research Square* DOI 10.21203/rs.3.rs-2924681/v1) as well as the ABD and C-terminal domains (Figure 1). LeuAC is thus homologous in that respect to the corresponding urzyme excerpted from *Geobacillus stearothermophilus*, TrpAC (6,14). However, the mass deleted from the catalytic domain (377 amino acids) is ~ five times more than were deleted (74 amino acids) from the TrpRS catalytic domain. That missing mass includes insertion modules that others have suggested were essential to aminoacylation (15). LeuAC required a modest number of sequence changes on the surface to stabilize the removal of the Connecting Peptide. Supplementary Figure S1 in the Supplementary Information for (12) summarizes these sequence changes.

Two consensus sequence motifs present in the Rossmann fold (HIGH and KMSKS) have been implicated multiple ways in catalysis of both amino acid activation and acyl-transfer to tRNA in many Class I aaRS. Their conservation across the superfamily has been an evolutionary curiosity, suggesting that they are central to catalysis. However, both histidine and lysine require complex biosynthetic pathways, and thus are unlikely to have been available for the earliest stages of codon-dependent translation (16–19).

The unexpectedly modest impact of mutating lysine residues to alanine in the LeuAC urzyme KMSKS signature (12) motivated us to examine the catalytic roles of both signatures in both LeuRS and LeuAC. An obvious way to do this is with a factorial design in which both histidines in the HxGH signature are also mutated to alanine separately and in combination. That combinatorial mutagenesis measures the energetic coupling between the two signatures. To our knowledge, and despite a landmark combinatorial mutagenesis of the KFGKT sequence in TyrRS (20–23), no one has examined the coordinated behavior of the two Class I catalytic signatures, even in a full-length Class I aaRS. Comparing that coupling in urzyme and full-length enzyme maps changes in the evolutionary time domain, hence measures the catalytic gain of function from integrating the domains deleted in the urzyme.

The comparison reinforces previous suggestions (24–27) that full-length Class I aaRS benefit from a significant intramolecular cooperativity that is absent from the corresponding urzymes. We identify side chain packing interactions between the hydrophobic side chains in each catalytic signature and a conserved cluster of nonpolar side chains in the anticodon-binding domain. Embedding the valine and methionine side chains of the LeuRS HVGH and KMSKS sequences into that cluster provides a structural rationale for their coordination during the catalytic cycle, consistent with their catalytic coupling in the full-length enzyme.

The anti-coupling between HVGH and KMSKS signatures in the urzyme has important implications for the emergence and early evolution of polypeptide catalysts. Surprisingly, the AVGA variant of the LeuAC urzyme benefits from both the wild type KMSKS signature and unfavorable coupling energy; it is thus the most active LeuAC variant and is actually more active at aminoacylation than the corresponding mutant of full-length LeuRS. Catalysis by aaRS urzymes does not require specific amino acid side chains that are highly conserved in the active sites of all Class I aaRS, thus significantly broadening the sequence space for

primordial catalysts and creating substantive experimental support for Wong's proposal that the genetic code coevolved with the amino acid biosynthetic pathways (16,17).

Finally, the consistent response to combinatorial mutational changes evident from the thermodynamic cycle analysis, and provision of a structural rationale for differential intramolecular coupling in the full-length enzyme both also decisively confirm the authenticity of catalysis by aaRS urzymes.

## MATERIALS AND METHODS

### 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 h then harvested and stored overnight at –20°C. The cell pellet was resuspended in 1× Ni-NTA buffer (20 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 mM β-ME) plus cComplete protease inhibitor (Roche) and lysed by three 15k psi passes on an Avestin Emulsiflex. Cell debris was pelleted at 4°C 30 min 20k rpm. The soluble fraction was heated at 80°C for 30 min 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 1× 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, brought to 50% 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 buffer (20 mM Tris, pH 7.4, 1 mM EDTA, 5 mM β-ME, 17.5% glycerol, 0.1% NP40, 33 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.25% glycine, 300 mM guanidine hydrochloride) plus cComplete protease inhibitor (Roche). LeuAC crude extract was then pelleted at 4°C 30 min 15k rpm to remove insoluble material. The extract was then 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 protein were concentrated, brought 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 samples cleaved by tobacco etch virus (TEV) protease, purified as described (28). Purity and cleavage efficiency was determined by running samples on PROTEAN® TGX (Bio-RAD) gels.

### Single turnover active-site titration assays

Active-site titration assays were performed as described (29,30). 3 μM of protein was added to 1x reaction mix

(50 mM HEPES, pH 7.5, 10 mM MgCl<sub>2</sub>, 5 μM ATP, 50 mM amino acid, 1 mM DTT, inorganic pyrophosphatase, and α-labeled [<sup>32</sup>P] ATP) to start the reaction at 37°C for LeuAC and on ice, or slightly above 0°C for LeuRS. The α labeling position allowed us to follow time courses of ADP and AMP production, as well as for ATP consumption. 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). ImageJ (31) 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 JMP16PRO™ (SAS Institute, Cary NC) to equation (1) (29):

$$\text{Product (calc)} = A * \exp(-k_{\text{chem}}^* \text{seconds}) - k_{\text{cat}}^* \text{seconds} + C \quad (1)$$

where  $k_{\text{chem}}$  is the first-order rate constant,  $k_{\text{cat}}$  is the rate of turnover,  $A$  is the amplitude of the first-order process or the size of the ‘burst’, and  $C$  is an offset. The burst size is proportional to the number of active molecules responsible for producing a product in the first round of catalysis in the added catalyst. Eq. (1) was also used for convenience to fit the appearance of nucleotide products, AMP and ADP. In that case the roles of  $A$  and  $C$  are inverted,  $C$  being the burst size and  $A$  the offset and both rate constants have the opposite sign.

### Burst size estimation from fitted parameters in equation (1)

For [ATP] decay curves, the fitted  $A$  value estimates  $n$  directly as  $n = A * [\text{ATP}] / [\text{Enzyme}]$ .  $C$  gives  $n = (1 - C) * [\text{ATP}] / [\text{enzyme}]$ . For exponentially increasing concentrations of product, the situation reverses, and  $n = (1 - A) * [\text{ATP}] / [\text{enzyme}]$  and  $n = C * [\text{ATP}] / [\text{enzyme}]$ . The very small variance of multiple estimates and the internal self-consistency discussed in (RESULTS:Single turnover...) justify these approximations.

### tRNA<sup>Leu</sup> aminoacylation assays

A plasmid encoding the *P. horikoshii* tRNA<sup>Leu</sup> (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.

We determined the active fraction of tRNA<sup>Leu</sup> by following extended acylation assays using full-length *P. horikoshii*

LeuRS until they reached a plateau. That plateau value was 0.38, which we used to compute tRNA<sup>Leu</sup> concentrations in assays with both LeuRS and LeuAC.

Aminoacylations were performed in 50 mM HEPES, pH 7.5, 10 mM MgCl<sub>2</sub>, 20 mM KCl, 5 mM DTT with indicated amounts of ATP and amino acids. Desired amounts of unlabeled tRNA—mixed with [α<sup>32</sup>P] A76-labeled tRNA for assays by LeuAC—were heated in 30 mM HEPES, pH 7.5, 30 mM KCl to 90°C for 2 min. The tRNA was then cooled linearly (drop 1°C/30 s) until it reached 80°C when MgCl<sub>2</sub> was added to a final concentration of 10 mM. The tRNA continued to cool linearly until it reached 20°C.

### Data processing and statistical analysis

Phosphorimaging screens of TLC plates were densitometered using ImageJ (31). Data were transferred to JMP16PRO™ via Microsoft Excel (version 16.49), after intermediate calculations. We fitted active-site titration curves to Eq. (1) using the JMP16PRO™ nonlinear fitting module.  $R^2$  values were all >0.97 and most were >0.99.

Factorial design matrices in Supplementary Tables S1 and S2 were processed using the Fit model multiple regression analysis module of JMP16PRO™ Pro, using an appropriate form of equation (2) (32)

$$Y_{\text{obs}} = \beta_0 + \sum \beta_i^* P_i + \sum \beta_{ij}^* P_i^* P_j + \varepsilon \quad (2)$$

where  $Y_{\text{obs}}$  is a dependent variable, usually an experimental observation,  $\beta_0$  is a constant derived from the average value of  $Y_{\text{obs}}$ ,  $\beta_i$  and  $\beta_{ij}$  are coefficients to be fitted,  $P_{ij}$  are independent predictor variables from the design matrix, and  $\varepsilon$  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  $\Delta G^\ddagger k_{\text{chem}}$ .

Multiple regression analyses of factorial designs exploit the replication inherent in the full collection of experiments to estimate experimental variances on the basis of  $t$ -test  $P$ -values, in contrast to the presenting error bars showing the variance of individual datapoints. Analyses reported here also entail triplet experimental replicates, which improve the associated analysis of variance.

## RESULTS

The experimental design matrix for this work is the 2<sup>3</sup> factorial design that tests all possible combinations of HVGH → AVGA and KMSKS → AMSAS mutations in both full-length LeuRS and the LeuAC urzyme. The eight variants were constructed, expressed, and purified as described previously (12) and in MATERIALS AND METHODS. We assayed all variants for leucine activation by single turnover active-site titrations (Supplementary Table S1) and for aminoacylation of tRNA<sup>Leu</sup> in triplicate (Supplementary Table S3). Active-site titration experiments provided suitable estimates of burst sizes for all variants of both LeuAC and LeuRS. However, first-order rate constants were too fast at room temperature to fit reliably for LeuRS variants. However, reactions on ice (at ~0°C) were

sufficiently slow to permit corresponding estimation of the mutational impact on leucine activation by LeuRS variants. The first section describes those results.

We then describe aminoacylation rates measured at saturating concentrations of leucine (12) and the highest achievable tRNA<sup>Leu</sup> concentrations. The next section compares mutational impacts on leucine activation and tRNA<sup>Leu</sup> aminoacylation by both LeuAC and LeuRS. A subsequent section places those results into the context of the evolutionary gain of function in evolving LeuRS from LeuAC. Finally, we provide a structural rationale for the enhanced energetic coupling between the two signatures in the fully-evolved enzyme.

### Single turnover experiments give consistent first-order rates for changes in ATP, ADP, and AMP for both LeuAC and LeuRS

Single turnover kinetic experiments, often called active-site titrations (AST) use substrate level amounts of enzyme. They measure the size of the burst corresponding to the first round of catalysis and are therefore the method of choice to estimate fractional activities of enzyme variants (29,33) and normalizing enzyme concentrations when computing  $k_{\text{cat}}$  values.

AST assays performed here to estimate the active fractions in all variants provide supplemental evidence that mutational impacts are comparable in both leucine activation and aminoacylation by LeuAC. We uncovered an unexpected phenomenon when we followed the time courses for all three adenine nucleotides in AST assays of aaRS urzymes (Figure 2). ATP consumption is accompanied by ADP production (12), in addition to the expected products, aminoacyl-5'AMP and AMP, which is produced by product release and hydrolysis. Prior to analyzing kinetic data for the thermodynamic cycles presented here, we studied these curves in detail to assure a consistent interpretation.

Quantitation of the LeuAC active-site titration curves for the three adenine nucleotides (Figure 2A) confirms first that they are additive;  $n$ -values (Figure 2B) from ATP consumption are quantitative sums of those for ADP and AMP produced (Figure 2C). LeuAC variant  $n$ -values from fitting  $C$  and  $A$  in Eq. (1) agree with a fractional error of  $\sim 0.09$ . ATP consumption  $\langle n_{\text{T}} \rangle$  values range from 0.64 to 0.69 for the different combinatorial variants.  $\langle n_{\text{D}} \rangle$  values (ADP) range from 0.49 to 0.57.  $\langle n_{\text{M}} \rangle$  values (AMP) range from 0.13 to 0.17. In each case,  $n_{\text{T}} = n_{\text{D}} + n_{\text{M}}$  within 2% (i.e. intercepts are  $\sim 0$ ; slopes are both  $\sim 1$ ). These data furnish a self-consistent picture of multiple catalyzed reactions occurring in the first round of catalysis (Supplementary Table S2).

LeuAC variant-specific column vectors in Figure 2B are quite strictly parallel; all six correlation coefficients have  $R^2$  values  $> 0.99$ . Thus, although combinatorial mutagenesis changes the active fractions, it has no effect on the relative magnitudes of the four different  $n$ -values. Strict linear dependence means that, for the purpose of normalizing the active fractions of different variants in analyzing aminoacylation rates, any of the  $n$ -values would be equally suitable. For the present purpose, we chose to normalize acylation rates using the  $n_{\text{T}}$  values, to estimate the minimum rate en-

hancements of acylation by full length LeuRS, relative to LeuAC.

The preceding analysis provides additional circumstantial evidence supporting our earlier proposal that ADP production arises from successive phosphorylations of bound leucyl-5'AMP. That proposal was originally based on two observations (12). (i) AMP production exhibits a first order rate early in the titration, followed by a sharp increase in the steady state rate (see Supplementary Figure S2(c) in (12)). (ii) The 3D structure of the LeuAC active-site configuration provides alternative ATP-binding sites that place the  $\gamma$ -phosphate group into suitable positions to phosphorylate the  $\alpha$ -phosphate of the bound adenylate first to ADP and then to ATP (see Figure 9 of (12)). If those two phosphorylation reactions regenerated ATP, it would promote bound Leu-5'AMP to a transition-state-like configuration. That would be consistent with the relatively independent burst sizes and relative proportions of ADP and AMP.

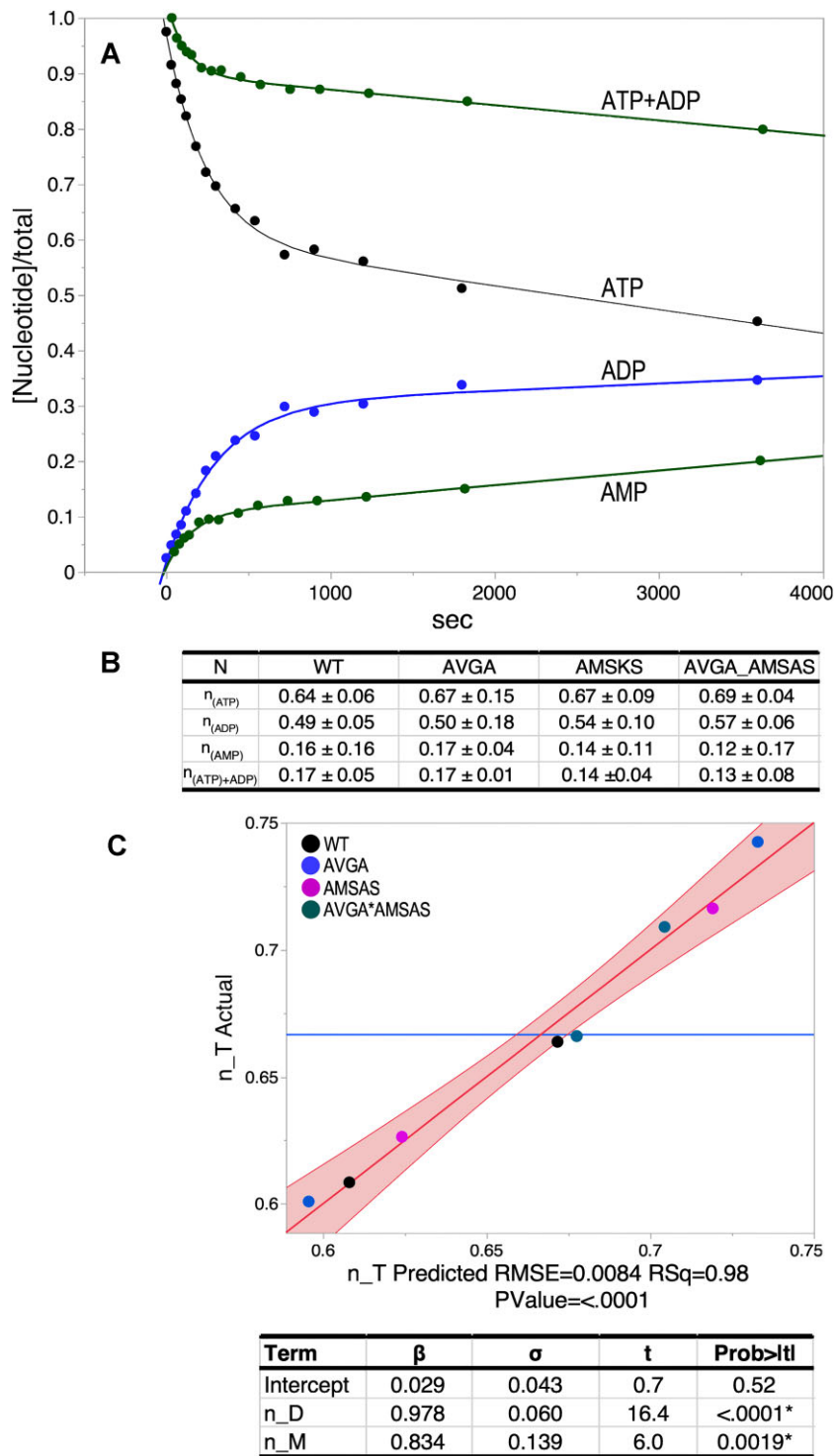
### Combinatorial mutants of HIGH and KMSKS catalytic signatures in LeuRS reduce its catalytic proficiency for tRNA<sup>Leu</sup> acylation to that of LeuAC and its mutational variants

Supplementary Table S4 summarizes the experimental data for aminoacylation of tRNA<sup>Leu</sup>. A histogram of transition-state stabilization free energies for tRNA<sup>Leu</sup> aminoacylation,  $\Delta G^\ddagger k_{\text{acyl}}$ , for all eight variants (Figure 3) reveals that the two catalytic signatures make profound contributions to the catalytic proficiency of the full-length enzyme. The distributions in the upper right of Figure 3B illustrate that  $\Delta G^\ddagger$  measurements for all LeuRS mutational variants have values comparable to those of the wild type Urzyme LeuAC and its corresponding mutations. Rather than forming a separate distribution in the neighborhood of the WT LeuRS, as observed for LeuAC variants, any mutation compromises all LeuRS variants to levels of the LeuAC urzyme and its combinatorial mutants. Active-site mutations therefore have severe effects on the full-length enzyme, but substantially smaller effects on the urzyme. Thus, neither wild-type urzyme active site nor any full-length LeuRS variant can achieve a comparable rate enhancement chemistry to that of the full-length enzyme.

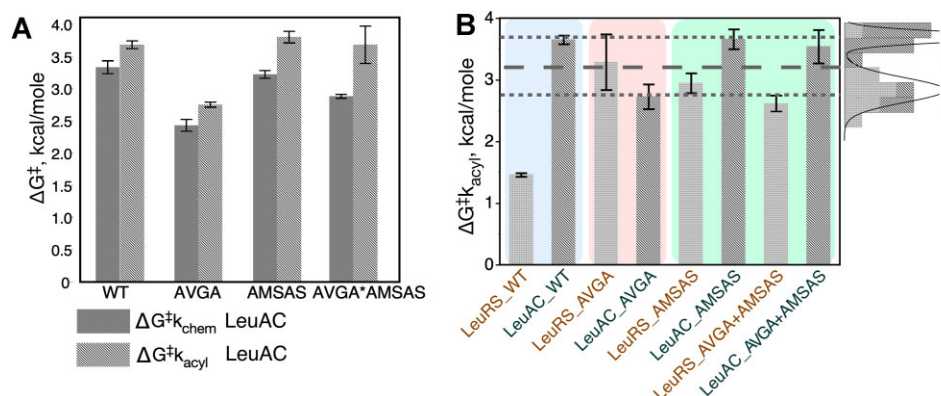
AST assays (Figure 2) furnish values of the internal first-order rate constant for first-round catalysis of leucine activation. Using these values, we can compare the impacts of mutation on the activation reaction. Figure 3A compares transition-state stabilization free energies of the four LeuAC variants in the two successive steps of tRNA<sup>Leu</sup> biosynthesis. Qualitatively, the main difference between them is that all variants have lower  $\Delta G^\ddagger$  for activation than for aminoacylation. The AVGA mutant is markedly faster than any of the others, and the AMSAS variant is slowest by a small margin over the double mutant. These differences are easier to appreciate when transformed by estimating parameters for the corresponding thermodynamic cycle [Eq. (2)].

### Thermodynamic cycle analysis emphasizes the gain of function resulting from ABD and CP domain acquisition

The factorial design matrices (Supplementary Tables S1, S2) enable us to attribute a magnitude and statistical signif-



**Figure 2.** N-value determination for the four LeuAC variants. (A) Single turnover time courses for the three adenine nucleotides. The green ATP + ADP curve represents the ATP consumption after accounting for ADP production and is almost a reflection of the (green) curve for AMP production. Thus, it likely represents *in situ* formation of leu-5' AMP. (B) Averaged n-values determined from the two physical constants, A (the amplitude of the burst) and C (the offset), with their standard deviations obtained for each variant by fitting to Eq. (1) as described in Materials and Methods. (C) Multivariate regression analysis of n-values averaged in B confirm that n-values for the appearance of ADP and AMP account almost exactly for ATP consumption. Solving the simultaneous equations  $[n_T = \sum_{i, \text{variants}} (\beta_D(C + n_D) + \beta_M(n_M) + \epsilon)]$  by least squares using JMP16PRO gave the table of values underneath the graph. Here, and elsewhere, the true value of the regression line has a 95% chance of falling within the red boundaries.



**Figure 3.** Individual transition-state stabilization free energies, in kcal/mole, for each LeuRS variant identified along the x axis. Error bars are two-way 95% confidence levels from triplicate determinations. (A) Transition stabilization free energies,  $\Delta G^\ddagger k_{\text{chem}}$  and  $\Delta G^\ddagger k_{\text{acyl}}$  represent relative rate accelerations. All LeuAC variants show comparable changes in amino acid activation and tRNA<sup>Leu</sup> aminoacylation. TS stabilizations are greater for activation than for acylation, as expected from differences in the corresponding uncatalyzed rates (59). (B) LeuRS variants are generally stronger catalysts than LeuAC variants, with the largest gap being between variants with wild type sequences (blue panel). AVGA variants have inverted rates, (rose panel) because the AVGA LeuAC variant has greater activity than the corresponding full-length LeuRS variant. Finally, AMSAS and double mutant variants have reduced activity in LeuAC and increased activity in LeuRS, relative to the corresponding AVGA variants (green panel). Distributions in the upper right show all 21 measurements for variants excepting WT LeuRS. Remarkably, the AVGA LeuAC mutant falls within the group of full-length variants that are more active than the other three LeuAC variants. Grey dashed lines identify a significant subset who's mean  $\Delta G^\ddagger k_{\text{acyl}}$ ,  $\sim 3.2$  kcal/mol, is significantly greater than that of WT LeuRS. Mutational variants of LeuRS, excepting AVGA, are nonetheless also better catalysts than those of corresponding LeuAC. This significant variation is difficult to interpret without thermodynamic cycle analysis.

importance of free energy contributions to transition-state stabilization for intrinsic effects of each signature sequence and the synergy between them. For this purpose we use a construct known as a thermodynamic cycle, advocated by Jencks (34) and popularized by Horowitz and Fersht (35). As originally formulated, construction of a thermodynamic cycle entails computing differences observed in activation free energies,  $\Delta(\Delta G^\ddagger)$ , at each corner of the implicit polygon of the factorial design. For a double mutant cycle, that involves a square with WT enzyme, the two single mutants, and the double mutant at the corners of a square (see Supplementary §II and Supplementary Figure S1).

We showed (26) that in practice it is more convenient to use multiple regression methods to estimate the coefficients in Eq. (2). When values of the experimental error,  $\epsilon$ , are small and the regression model explains a high ( $\geq 0.9$ ) portion of the variation in experimental data points, a thermodynamic cycle is equivalent to changing the coordinate system of the experimental free energies in Figure 3. Supplementary §II examines this equivalence in more detail, and Supplementary §III outlines how regression coefficients for main effects and lower-order interactions depend on where along the highest-order interaction they are evaluated.

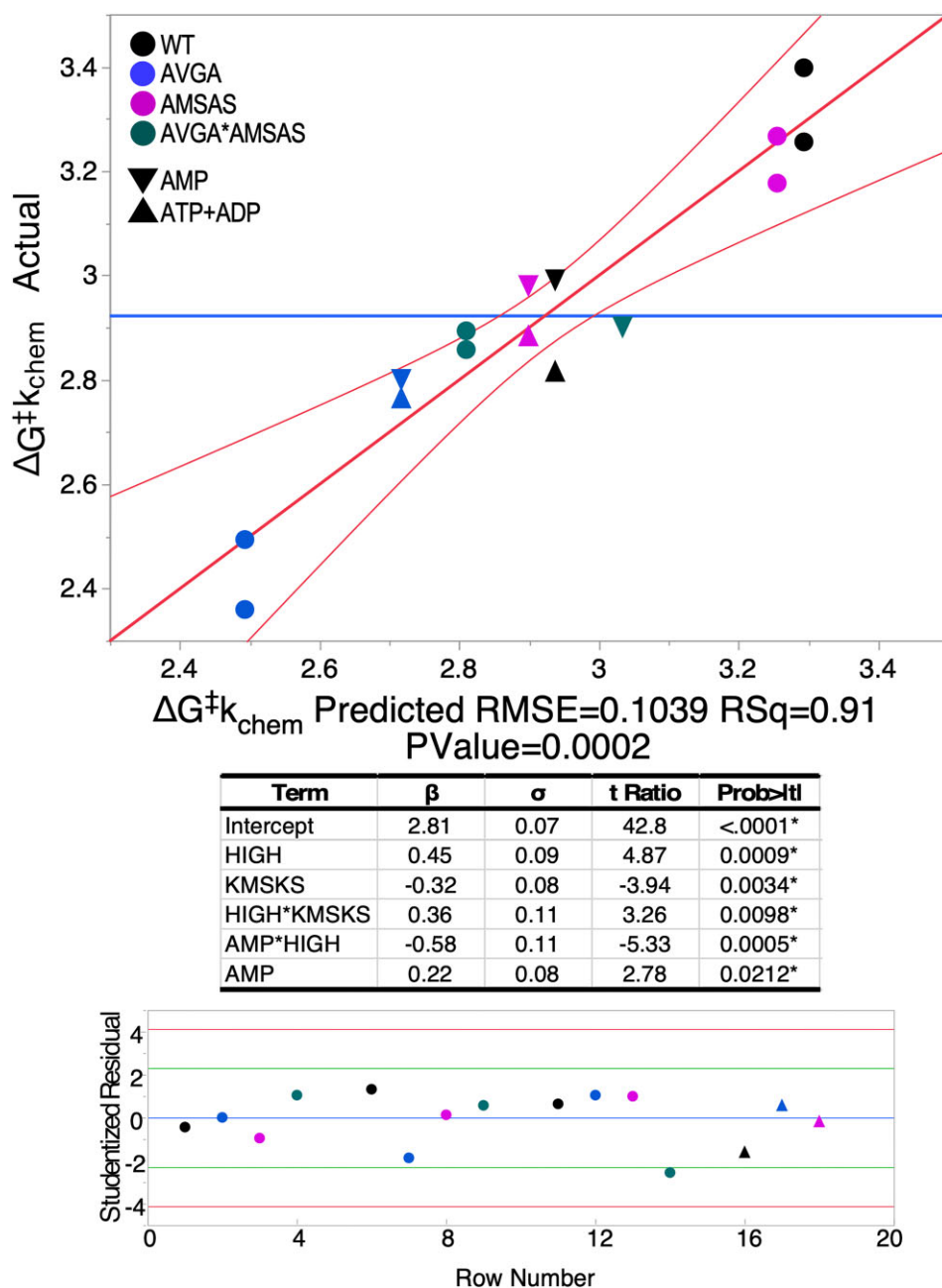
The regression model for the thermodynamic cycle for  $\Delta G^\ddagger k_{\text{chem}}$  from first-order rates, for all three nucleotides either consumed or produced in amino acid activation (Figure 4) reveals several notable details. First, the distribution of Studentized residuals associated with the contributing data points (the prediction error,  $\epsilon$ , divided by its adjusted standard error) lies between  $-4 \leq 4$ , so there are no outliers. Second, there are five significant predictors listed in the table. Their Student t-test probabilities are all  $\leq 0.02$ , hence are statistically significant. Third, because neither ATP nor ADP are significant predictors of the first-order rate, there is no significant difference between rates for ATP consump-

tion and ADP formation for the four variants. AMP production, on the other hand, is significantly different, and the AMP\*HVGH  $\beta$ -coefficient means that it depends on whether the HxGH sequence is native or mutant. We return to this point after considering mutational effects on tRNA<sup>Leu</sup> acylation.

First-order rate constants from AST,  $k_{\text{chem}}$ , are independent of enzyme concentrations, and so do not require normalization for active fraction before converting to free energies for multiple regression analysis to construct thermodynamic cycles. Figure 4 highlights the statistical quality of the regression model (Figure 5 omits the corresponding models). Figure 5 displays histograms of the  $\beta$ -coefficients for the acylation reaction in kcal/mol, with error bars denoting the 95% confidence levels. Arrows in Figure 5B form cycles showing that the net free energy change of the successive steps total zero. The difference between left and right vertical arrows must equal that between top and bottom arrows, and is equal to the free energy,  $\Delta(\Delta G^\ddagger)$  for the HVGH\*KMSKS interaction.

The most striking feature in Figure 4 is that samples at the extremes (the black dots = WT; the blue dots = AVGA) are opposite of what might be expected. The AVGA mutant is, by  $\sim 1$  kcal/mol, the most active catalyst in leucine activation and the WT variant is least active. Moreover, although the sign of the  $\beta_{\text{KMSKS}}$  coefficient is negative, signifying that the wild-type lysine residues favor catalysis, the synergy between it and the HIGH sequence,  $\beta_{\text{HVGH}^*\text{KMSKS}}$ , is positive. Thus, when it comes to stabilizing the transition state for the internally catalyzed reaction, the two signatures actually work against one another. That, and the positive  $\beta_{\text{HIGH}}$  value, account for the superiority of the AVGA variant, which is  $\sim 5$  times more active than wild-type LeuAC.

Figure 5 compares thermodynamic cycles (Figure 5B) for leucine activation (Figure 5A) and tRNA<sup>Leu</sup> aminoacy-

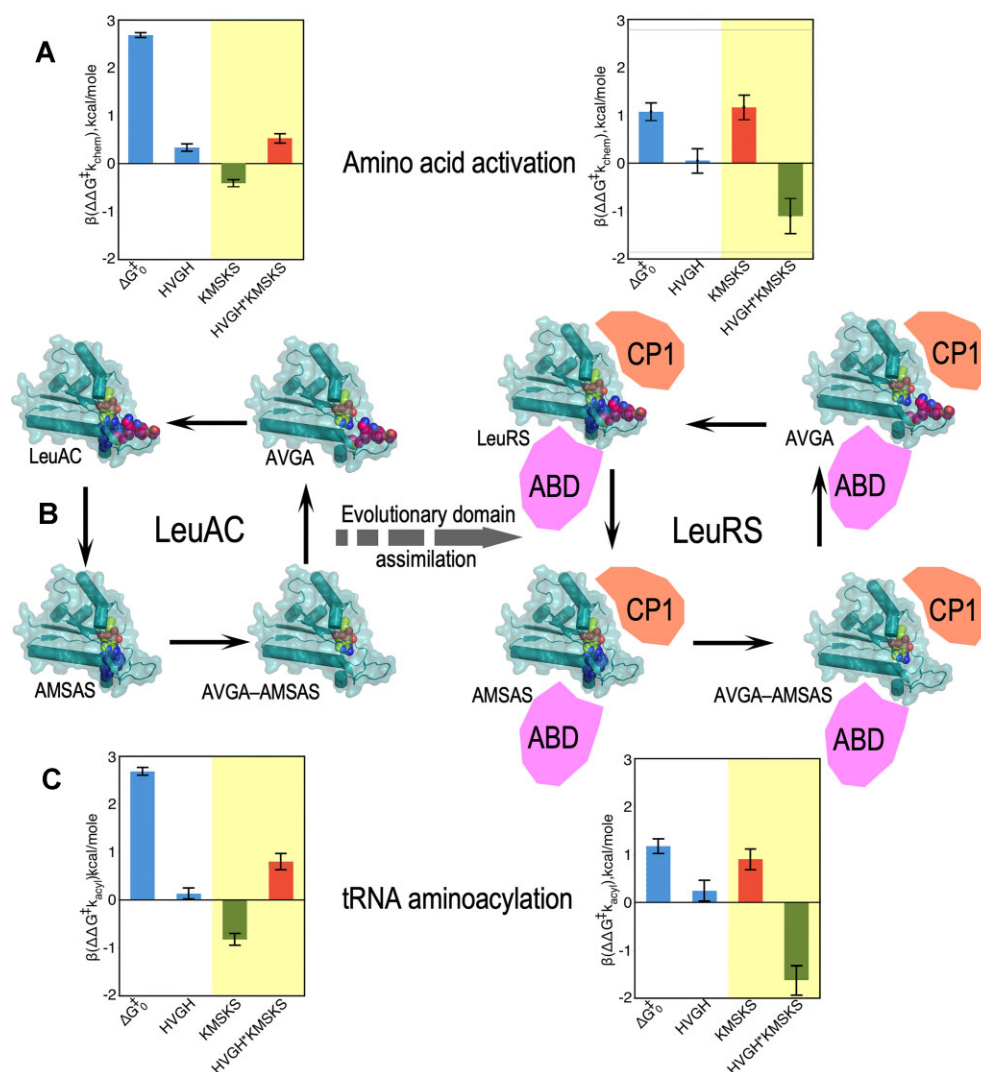


**Figure 4.** Multiple regression model for the activation free energies,  $\Delta G^\ddagger k_{\text{chem}}$  for the internal first-order rate constant,  $k_{\text{chem}}$ , for all LeuAC variants for time courses in Figure 1A. Circles represent ATP consumption and ADP production; triangles represent the two alternative measurements for AMP production.  $\beta$ -coefficients are in kcal/mole. The coloring scheme for variants is given in the upper left-hand corner. Studentized residuals ( $\varepsilon_{St}$ , bottom) are all  $-4.0 < \varepsilon_{St} < 4.0$  confirming the absence of outliers.

cylation (Figure 5C) by LeuAC and LeuRS.  $\Delta G^\ddagger_0$  is the mean activation free energy for all four variants of the respective catalysts; the remaining bars represent the additional transition-state stabilization ( $\beta < 0$ ) or destabilization ( $\beta > 0$ ) attributable to wild type HVGH, KMSKS, sequences and their two-way interaction or synergy.  $\beta$  coefficients for the main effects of the two mutations and their interaction tell us much about how the two catalytic signatures impact LeuAC catalysis. Wild-type residues in each location have opposite effects. Wild-type

lysine residues in the KMSKS signature enhance catalysis modestly ( $\Delta G^\ddagger_{\text{KMSKS}} = -0.32$  kcal/mole) compared to alanine mutants. Wild-type histidine residues in the HVGH signature actually reduce catalysis by a comparable amount (positive  $\Delta G^\ddagger_{\text{HVGH}} = 0.45$  kcal/mol). The interaction term is also positive ( $\Delta G^\ddagger_{\text{HVGH}*\text{KMSKS}} = 0.36$  kcal/mol), so the wild-type lysine residues in the KMSKS signature intensify the negative impact of the native HVGH sequence, and seemingly paradoxically, the presence of wild type KMSKS intensifies catalysis by mutant AVGA





**Figure 5.** Thermodynamic cycles for LeuAC (left) and LeuRS showing amino acid activation (top) and tRNA aminoacylation (bottom). Data in Figure 3B and Supplementary Table S3 were transformed via Eqn. (2) into multiple regression coefficients,  $(\beta_i, \beta_{ij})$ , to determine successive changes in activation free energy,  $(\Delta G^\ddagger k_{chem})$ . (A)  $\Delta G^\ddagger k_{acyl}$ , (C), for successive mutations shown by single-headed arrows around the square in (B) Error bars are derived from the estimated standard deviation of the  $(\beta_i, \beta_{ij})$  values along the x-axes in A and C. Intrinsic effects of LeuAC and LeuRS catalyts and their HVGH signatures are similar in sign and relative magnitude (white panels). KMSKS and two-way HVGH\*KMSKS  $\Delta(\Delta G^\ddagger)$  values however behave radically differently with and without the anticodon-binding (ABD) and CP domains (yellow panels). The two-way synergy between the two signatures (HVGH\*KMSKS) changes from anti-cooperative and unfavourable in LeuAC (red bar) to strongly cooperative and favourable (green bar) in LeuRS, increasing transition-state stabilization for activation by  $\sim 2$  kcal/mol and aminoacylation of  $tRNA^{Leu}$  by  $\sim 2.6$  kcal/mol. Cartoons suggest the acquisition of CP and ABD domains in full-length LeuRS. The thick dashed arrow represents our interpretation that the domain acquisition is responsible for in the different histograms from LeuAC to LeuRS, hence representing evolutionary time. Note that LeuRS  $k_{chem}$  values were measured on ice and underestimate the values at  $37^\circ\text{C}$ .

sequence. Combinatorial mutagenesis has comparable impacts on leucine activation and  $tRNA^{Leu}$  aminoacylation by LeuAC.

The KMSKS and HVGH\*KMSKS effects behave very differently in full-length LeuRS. The inversion of these effects from left to right represents catalytic consequences of adding the anticodon binding domain and CP (as well as CP2) to the LeuAC urzyme. That consequence is a substantial increase in transition-state stabilization for both reactions in LeuRS by enforcing coupled behavior on the two catalytic signatures, as discussed further in the next section and the DISCUSSION.

Thermodynamic cycle parameters built from  $\Delta G^\ddagger k_{chem}$  values from time courses of the three nucleotides plus acylation illustrate how mutations affect all four rate constants (Figure 6). The main effects of wild-type HVGH and KMSKS signatures and their two way interaction are essentially identical in the LeuAC urzyme (Figure 6A, C). Mutational changes also impact both steps of the reaction by full-length LeuRS consistently (Figure 6B, C). Those for AMP production and  $tRNA^{Leu}$  aminoacylation by LeuRS are nearly colinear (Figure 6B). However, an important distinction for LeuRS emerges in Figure 6C. Only AMP production by full-length LeuRS has the same sensitivity to

A LeuAC				
Effect	ATP	ADP	AMP	Acyl
Intercept	2.77	2.7	2.87	3.34
HIGH	0.53	0.73	0.15	0.52
KMSKS	-0.16	-0.19	-0.06	-0.41
(HIGH)*(KMSKS)	0.48	0.63	0.12	0.80

B LeuRS				
Effect	ATP	ADP	AMP	Acyl
Intercept	1.52	1.39	1.18	3.16
HIGH	-0.67	-0.41	-0.11	-0.41
KMSKS	0.42	0.91	-0.22	-0.75
(HIGH)*(KMSKS)	-1.44	-0.93	-0.79	-2.16

C LeuRS				
	T	D	M	Acyl
T		0.98	0.51	0.47
D	0.99		0.36	0.33
M	0.99	0.99		0.99
Acyl	0.99	0.99	0.93	

D LeuRS					
Variant	Active fraction	ADP/Tot	AMP/Tot	ADP %	AMP %
WT	0.84	0.19	0.66	0.22	0.79
AVGA	0.40	0.49*	0.03	1.23*	0.07
AMSAS	0.15	0.14	0.01	0.91	0.08
Double	0.16	0.49*	0.02	3.00*	0.15

**Figure 6.** Consistent mutational impact on leucine activation and tRNA<sup>Leu</sup> acylation. (A) Main effects and coupling energy between HVGH and KMSKS signatures for each of the reactions measured by AST and by tRNA<sup>Leu</sup> aminoacylation by LeuAC. (B) A similar table for effects on the same reactions by LeuRS. (C) Correlation coefficients between the four, three element column vectors (i.e. excluding the intercepts) for the attributed free energies in A and B. LeuRS correlations are in the upper triangular matrix; LeuAC values are in the lower triangular matrix. Blue denotes near identical vectors; red denotes non-colinear vectors. Supplementary Figure S6 provides a scatterplot matrix of the full set of these values. (D) Burst sizes for the three reactions measured by AST for full-length LeuRS. The last two columns give the percentages of the total ATP consumption as ADP and AMP. Note that wild type LeuRS produces mostly AMP, consistent with the correlation between AMP production and acylation in C.

mutations as does aminoacylation. In contrast, ATP consumption and ADP production exhibit similar mutational profiles that both respond favorably to the AMSAS mutation (i.e.  $\Delta G^\ddagger k_{\text{chem}} > 0$ ).

Products of the internal first-order reactions in the first round of catalysis also distribute differently in full-length LeuRS and the LeuAC urzyme. The ratio, M/D, of AMP to ADP produced (Supplementary Figure S4A;  $R^2 = 0.93$ ;  $P \leq 0.0001$ ) is inversely proportional to the presence of the native KMSKS sequence. Thus, increased ATP consumption associated with LeuAC variants lacking the wild-type lysine residues is also associated with lower amounts of the expected product, AMP. In contrast, the M/D ratio in full-length LeuRS variants depends fully on both main effects (HVGH and KMSKS) and on their two-way interaction (Supplementary Figure S4B;  $R^2 = 0.99$ ;  $P < 0.0001$ ), in a manner similar to the dependence of the activation free energies,  $\Delta G^\ddagger k_{\text{chem}}$ .

Figure 6C shows that contributions of the HVGH and KMSKS signatures and their two-way interactions to ATP consumption and ADP production by LeuRS are distinctly different from those contributing to AMP production and tRNA aminoacylation. In fact, their mutational profiles are closely anticorrelated to those for AMP production and acylation, as well as those for all four reactions by LeuAC ( $\langle R^2 \rangle = 0.85 \pm 0.1$  see Supplementary Figure S5). Together with thermodynamic cycles for rates of AMP production by LeuAC and LeuRS (Supplementary Figure S3), these observations suggest an overriding distinction between catalysis of aminoacylation by full-length LeuRS and all 7 remaining variants illustrated in Figure 3B. Deletion of any contributor to the wild-type full-length LeuRS mechanism—either of the various domains missing in the urzyme or of either or both signature sequence—changes the catalytic mechanism in similar underlying manner (see again the distribution in the upper right of Figure 3B).

### Nonpolar sidechains V51 and M651 anchor both HVGH and KMSKS signatures securely into the ABD

It is worth noting that although only minimal structural data are available for any aaRS urzymes (36), neither of

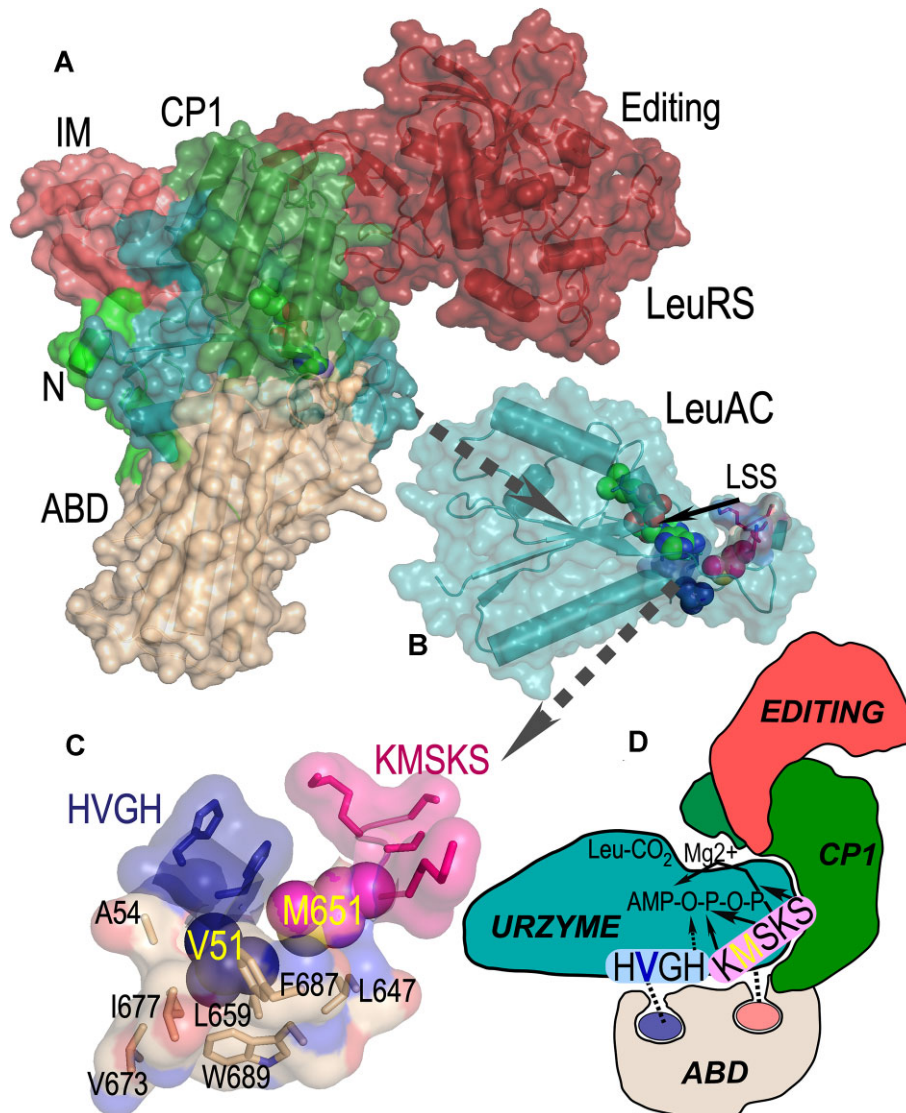
the hydrophobic residues in the two signature sequences appears from crystal structures of full-length Class I aaRS to be engaged in any significant nonpolar packing interactions within in LeuAC itself (Figure 7B). By contrast, a well-developed packing network of nonpolar amino acids including the two hydrophobic side chains together with residues from the ABD in full-length LeuRS anchors the valine and methionine residues within that domain (Figure 7). Delaunay tessellation and likelihood scoring (37,38) identified a homologous network in TrpRS (39), and was used here to identify residues detailed in Figure 7C. Similar networks can be identified in all full-length Class I aaRS. They appear to be both necessary and sufficient to coordinate the behavior of the two catalytic signatures, as shown schematically in Figure 7D.

## DISCUSSION

The role of cooperative interactions between active-site residues in enzyme catalysis is a recurring question that is seldom addressed directly by experimentation (40), despite well-established protocols for thermodynamic cycle analysis of combinatorial mutants (35). The thermodynamic cycle is a revealing linear transformation of the experimental transition state stabilization free energies,  $\Delta G^\ddagger$ , which as noted elsewhere are additive. The alternative coordinate system highlights the magnitude and sign of individual and energetically coupled contributions to catalysis.

### Use of <sup>32</sup>P $\alpha$ -ATP in active site titrations reveals hidden mechanistic microheterogeneity

Introducing the use of <sup>32</sup>P $\alpha$ -ATP in active site titration assays has expanded our appreciation of the complexity of first-order events in the first round of catalysis. The ability to track all three adenine nucleotides reveals unexpected production of small amounts of ADP even by full length LeuRS and especially in LeuRS active-site variants (Supplementary Table S2). Indeed, as discussed in RESULTS in other contexts, loss either of the additional domains or the intact catalytic apparatus in the active site is sufficient to



**Figure 7.** Internal mechanics of the LeuRS (B) and LeuAC (C) active sites. (A) Space-filling cartoon of intact LeuRS (PDB ID 1WZ2) showing the same domains with the same color scheme as in Fig. 1A. (B) LeuAC urzyme, LSS is a leucyl-5'-sufoamyl adenosine analog of the activated amino acyl adenylate. (C) The 'Enforcer'. Nonpolar side chain packing anchors V51 (blue) from the HVGH (blue) and M651 (magenta) from the KMSKS signatures firmly within the ABD. Illustrated residues participate in high statistical potential Delaunay tetrahedra surrounding the nonpolar side chains of the two signature sequences. (D) Cartoon highlighting the tight mechanical linkage of the two signature sequences to the ABD, and consequently to their cooperative motions relative to the urzyme.

dramatically alter the flow of chemical free energy during amino acid activation.

The equivalence between ATP consumption and the sum of ADP and AMP burst sizes, combined with the similarity of first-order rates for the three processes by LeuAC implicate similar catalytic machinery in both ADP and AMP production. The KMSKS sequence in full-length Class I aaRS is involved with transient binding of the PP<sub>i</sub> leaving group of ATP (see, for example, the effects of removing PP<sub>i</sub> from the free energy surface traversed during catalysis by TrpRS (41)). A more extensive discussion of this question is in Supplementary §IV.

#### Emergent coupling deepens evidence for the authenticity of urzyme catalysis

We cannot overlook the obvious conclusion that the contrast between the uncoupled behavior of the two catalytic signatures in LeuAC and their strong coupling by the ABD and CP in full length LeuRS reinforces the authenticity of both LeuAC (12) and TrpAC (25) catalysis. The evolution of intramolecular coupling shown here therefore materially strengthens the argument that aaRS urzymes represent valid experimental models for the ancestral assignment catalysts that originally enabled nature to mine the immense functional diversity represented by proteins (7).

### The LeuAC studies complement and extend earlier thermodynamic cycle analyses of TyrRS and TrpRS mechanisms

In the present case the coupling between catalytic signatures might, in principle, arise solely from the sequence differences far from the active site on the newly created molecular surface of LeuAC, rather than from the removal of the CP and anticodon-binding domains. We cannot envision any practical way to rule this possibility out, as the native sequences lead to insoluble aggregates of LeuAC. Prior studies comparing TrpAC with TrpRS substantially strengthen confidence in the interpretation outlined here, which implicate domain movement and intramolecular communication in the enhanced function of the full-length enzyme.

Class I aminoacyl-tRNA synthetase enzymes (aaRS) afford several examples of how thermodynamic cycle analysis probes structure-function relationships. First and Ferstl showed that the effects of mutating residues in the KMSKS loop were localized quite specifically along the reaction profile to destabilize the ground state pre-transition state and stabilize the transition state for tyrosine activation by TyrRS (20). The KMSKS signature breaks the synergistic binding of amino acid and ATP in the pre-transition state (20–23).

Destabilization by the KMSKS signature of the transition states for both activation and aminoacylation in LeuRS appears to be inconsistent with that observation. A plausible rationale for this apparent discrepancy is that the behavior in TyrRS also depends on interaction between the two signatures. That interpretation is especially appealing because LeuAC exhibits the opposite dependence: the wild-type KMSKS sequence itself destabilizes both transition states (Figure 5). Its catalytic contribution to both reactions becomes favorable only by virtue of the two-way interaction.

A packing motif (42) called the D1 switch (39) links the two  $\beta$ -strands and  $\alpha$ -helix of the first crossover connection to the N-terminal  $\beta$ -strand of the second crossover connection of the Rossmann dinucleotide-binding domain, creating a conformational transition state that mediates shear during relative domain motion of the ABD and CP and imposing discrete pre-transition and products conformations in TrpRS (41,43,44) and likely in other Class I aaRS. Combinatorial mutagenesis of four residues from the D1 switch, assayed with both  $Mg^{2+}$  and  $Mn^{2+}$  established a five-dimensional thermodynamic cycle. That cycle revealed that the packing motif, which is  $\sim 20$  Å from the active-site metal, is nevertheless coupled by  $-6.2$  kcal/mol to the catalytic function of the metal (26). That contribution nearly equals the full catalytic contribution,  $-6.4$  kcal/mol, of the metal measured by assaying metal-free TrpRS (45). A complementary, modular thermodynamic cycle consisting of TrpRS, its urzyme, TrpAC and TrpAC plus either the CP insert (i.e. the intact catalytic domain) or the ABD showed that both CP and the ABD alone actually reduced the activity of TrpAC in aminoacylation of tRNA<sup>Trp</sup>, but that both together restored full activity, a contribution of  $\sim 5$  kcal/mol—nearly the entire contribution of  $Mg^{2+}$  (25). Finally, both the combinatorial mutagenesis (24) and modular cycles (25) also impli-

cate the corresponding energetic coupling in amino acid specificity.

Coupling energies between D1 switch residues and  $Mg^{2+}$  in transition-state stabilization are comparable to those between the TrpRS ABD and CP domains (24–27). The minimum action path connecting successive structures along the TrpRS structural reaction profile implies that repacking D1 switch residues in the conformational transition state is rate-limiting between pre-transition state and products domain configurations (41,43,44). These fragmentary experimental vignettes consistently suggest that a central function of the ABD and CP domains is to impose coordinated motion on components of the active site, such that they all assemble simultaneously into a catalytically active, discriminating configuration (24,25).

### Experimental exploration of the time domain

The thermodynamic cycles comparing LeuRS to its putative ancestral urzyme form in Figures 3–5 extend experimentation into the evolutionary time domain (46–53). In this manner we identify quantitatively the explicit gain of function induced in Class I leucyl-tRNA synthetase by the acquisition of the connecting peptide 1 (CP) insertion and anticodon-binding (ABD) domains. The new data reported here reinforce the picture outlined in previous paragraphs. In LeuAC, the absence of the two larger domains leaves the two catalytic signatures uncoordinated, contributing in unexpected ways to transition state stabilization. In full length LeuRS, on the other hand, although neither signature alone can enhance transition-state stabilization, their combination, in the context of the two additional domains, furnishes substantial catalytically productive synergy. That we now have identified much the same phenomena we observed for TrpRS, the smallest Class I aaRS in one of the largest as well as the smallest, Class I aaRS suggests generality in the superfamily. The superfamily-wide conservation of the ‘enforcer’ packing motif illustrated in Figure 7C, in which high SNAPP potential (37,38) hydrophobic clusters anchor the sole hydrophobic side chains of each catalytic signature, coordinating their motion reinforces its generality.

### Successively acquired CP and anticodon-binding domains may have had distinct, less obvious selective advantages

We have argued that aaRS of both Classes began as 46-residue polypeptides called ‘protozymes’ that accelerated the rate of amino acid activation  $\sim 10^6$ -fold (54). That hypothesis gained support from validation by Tamura’s group (55). Thus, it is possible that the HxGH signature has roots in the protozyme. Class I aaRS urzymes, however, are more than twice as big ( $\sim 130$  residues), and the KMSKS signature must have roots in that subsequent stage. The selective advantage of the KMSKS signature likely contributed to the acquisition of the second half of urzymes, which accelerated the rate of activation by  $10^3$ -fold over that for the protozyme.

As noted previously (25), neither CP nor the ABD enhanced either specificity or catalytic rate enhancement of

the TrpRS urzyme. This prompts the question of how nature selected either domain except in the unlikely case that the urzyme acquired both simultaneously. The AST assays performed to assess the fraction of active catalysts in each sample suggest a possible resolution of that conundrum. The proportion of total ATP usefully converted to AMP production by LeuAC (0.21) is roughly 25% that (0.8; Figure 6D) observed for LeuRS. If the CP domain stabilizes closed, productive forms of the urzyme so that it used ATP more efficiently, that would certainly have given it a selective advantage, even without imposing the coupling between the two signatures. This hypothesis is experimentally testable.

### Significant rate acceleration by early polypeptide catalysts did not require sophisticated amino acid side chains

The LeuAC AVGA variant ( $\langle \Delta G^\ddagger \rangle = 1.9$  kcal/mol) is quantitatively more active than the same variant of the full-length LeuRS ( $\langle \Delta G^\ddagger \rangle = 2.1$  kcal/mol). That brings us to what is perhaps the most consequential implication from the data in Figures 2–4. The overall rate enhancements observed for amino acid activation by aaRS urzymes excerpted from two Class I (TrpRS (6,7,14), LeuRS (12)) and one Class II aaRS (HisRS (10,25)) represent ~60% of the overall transition-state stabilization free energy observed for the corresponding full length aaRS. How much of that must we attribute to the precise orientation of specific amino acid side chains in the active site?

Our data suggest that the surprising answer to that question is: apparently very little, if any of it. Indeed, the superiority of the AVGA LeuAC mutant means that the two histidine residues in that highly conserved signature actually reduce the rate of aminoacylation unless they are coupled to the KMSKS sequence by the additional domains. Moreover, the AMSAS mutant is almost as active as the WT LeuAC. These observations are consistent with the previous mutational analysis of the TrpAC urzyme, where mutation of an aspartate at the N-terminus of the C-terminal  $\alpha$ -helix of the second crossover connection in the Rossmann fold reduces activity by ~200-fold in TyrRS but increases activity ~25-fold in the TrpRS urzyme.

We previously noted that secondary structural; i.e. backbone interactions in Class I and II urzymes could account for tRNA (56,57) and amino acid (58) substrate selection (7). Together with our previous evidence that the TrpAC urzyme may be a catalytically active molten globule (36), these results substantiate the possibility that substantial catalytic rate enhancement does not require precise orientation of specialized amino acid side chains. It now seems that primordial protein catalysts achieved high rate enhancements and a modicum of substrate specificity largely via physical properties of the active-site pocket that depend largely on backbone interactions that do not require specific amino acid side chains.

That conclusion lends a strong experimental basis for Wong's coevolution theory for the emergence of the genetic code and its coevolution with metabolic pathways necessary to synthesize amino acids that were

not produced in abundance by prebiotic geochemistry (16,17).

### DATA AVAILABILITY

JMP is a product of the SAS corporation, Research Triangle, NC, USA. All rate data are provided in Supplementary Tables, are accessible from The Protein Databank, or upon request from the author.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

P. Wills provided continual constructive discussions. E. First and L. Betts made helpful comments on the manuscript.

*Author contributions:* J.J.H.E. and Q.C.T. designed the combinatorial LeuAC mutants and performed all assays. Q.C.T. constructed and assayed the combinatorial LeuRS mutants. C.W.C., J.J.H.E., and G.C.T. all performed data analysis. J.D. provided phylogenetic insight for Figure 1s as well as the evolutionary history of the signature sequences. C.W.C. wrote the manuscript, and all authors contributed to and approved the final figures and text.

### FUNDING

Alfred P. Sloan Foundation Matter-to-Life program [G-2021-16944]. Funding for open access charge: Alfred P. Sloan Foundation [G-2021-16944].

*Conflict of interest statement.* None declared.

### REFERENCES

- Rubio Gomez, M.A. and Ibba, M. (2020) Aminoacyl-tRNA synthetases. *RNA*, **26**, 910–936.
- Terada, T., Nureki, O., Ishitani, R., Ambrogelly, A., Ibba, M., Söll, D. and Yokoyama, S. (2002) Functional convergence of two lysyl-tRNA synthetases with unrelated topologies. *Nat. Str. Biol.*, **9**, 257–262.
- Burbaum, J.J. and Schimmel, P. (1991) Assembly of a class I tRNA synthetase from products of an artificially split gene. *Biochem.*, **30**, 319–324.
- Burbaum, J.J., Starzyk, R.M. and Schimmel, P. (1990) Understanding structural relationships in proteins of unsolved three-dimensional structure. *PROTEINS: Struct. Funct. Gen.*, **7**, 99–111.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, **347**, 203–206.
- 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.
- Carter, C.W. Jr and Wills, P.R. (2021) The roots of genetic coding in aminoacyl-tRNA synthetase duality. *Annu. Rev. Biochem.*, **90**, 349–373.
- Carter, C.W. Jr (2017) Coding of class I and II aminoacyl-tRNA synthetases. *Adv. Exp. Med. Biol.*, **18**, 103–148.
- Carter, C.W. Jr (2014) Urzymology: experimental access to a key transition in the appearance of enzymes. *J. Biol. Chem.*, **289**, 30213–30220.
- 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.

11. Fukunaga, R. and Yokoyama, S. (2005) Crystal structure of leucyl-tRNA synthetase from the archaeon *Pyrococcus horikoshii* reveals a novel editing domain orientation. *J. Mol. Biol.*, **346**, 57–71.
12. Hobson, J.J., Li, Z. and Carter, C.W. Jr (2022) A leucyl-tRNA synthetase urzyme: authenticity of tRNA synthetase urzyme catalytic activities and production of a non-canonical product. *Int. J. Mol. Sci.*, **23**, 4229.
13. Cusack, S., Yaremchuk, A. and Tukul, M. (2000) The 2 Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J.*, **19**, 2351–2361.
14. Pham, Y., Li, L., Kim, A., Erdogan, O., 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.
15. Zhou, X.L., Zhu, B. and Wang, E.D. (2008) The cp2 domain of leucyl-tRNA synthetase is crucial for amino acid activation and post-transfer editing. *J. Biol. Chem.*, **283**, 36608–36616.
16. Wong, J.T.-F., Ng, S.-K., Mat, W.-K., Hu, T. and Xue, H. (2016) Coevolution theory of the genetic code at age forty: pathway to translation and synthetic life. *Life*, **6**, 12.
17. Wong, J.T.-F. (2005) Coevolution theory of the genetic code at age thirty. *Bioessays*, **27**, 416–425.
18. Bada, J.L. (2013) New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chem. Soc. Rev.*, **42**, 2186–2196.
19. Miller, S.L. (1953) A production of amino acids under possible primitive earth conditions. *Science*, **117**, 528–529.
20. First, E.A. and Fersht, A.R. (1995) Analysis of the role of the KMSKS loop in the catalytic mechanism of the tyrosyl-tRNA synthetase using multimeric cycles. *Biochem.*, **34**, 5030–5043.
21. First, E.A. and Fersht, A.R. (1993) Involvement of threonine 234 in catalysis of tyrosyl adenylate formation by tyrosyl-tRNA synthetase. *Biochem.*, **32**, 13644–13650.
22. First, E.A. and Fersht, A.R. (1993) Mutation of lysine 233 to alanine introduces positive cooperativity into tyrosyl-tRNA synthetase. *Biochem.*, **32**, 13651–13657.
23. First, E.A. and Fersht, A.R. (1993) Mutational and kinetic analysis of a mobile loop in tyrosyl-tRNA synthetase. *Biochem.*, **32**, 13658–13663.
24. Weinreb, V., Li, L., Chandrasekaran, S.N., Koehl, P., Delarue, M. and Carter, C.W. Jr (2014) Enhanced amino acid selection in fully-evolved tryptophanyl-tRNA synthetase, relative to its urzyme, requires domain movement sensed by the D1 switch, a remote, dynamic packing motif. *J. Biol. Chem.*, **289**, 4367–4376.
25. Li, L. and Carter, C.W. Jr (2013) Full implementation of the genetic code by tryptophanyl-tRNA synthetase requires intermodular coupling. *J. Biol. Chem.*, **288**, 34736–34745.
26. Weinreb, V., Li, L. and Carter, C.W. Jr (2012) A master switch couples Mg<sup>2+</sup>-assisted catalysis to domain motion in *B. stearothermophilus* tryptophanyl-tRNA synthetase. *Structure*, **20**, 128–138.
27. Budiman, M., Knaggs, M.H., Fetrow, J.S. and Alexander, R.W. (2007) Using molecular dynamics to map interaction networks in an aminoacyl-tRNA synthetase. *Proteins Struct. Funct. Bioinform.*, **68**, 670–689.
28. Parks, T.D., Leuther, K.K., Howard, E.D., Johnston, S.A. and Dougherty, W. (1994) Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.*, **216**, 413–417.
29. 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 aminoacyl-tRNA synthetases. *Biochem.*, **14**, 1–4.
30. Francklyn, C.S., First, E.A., Perona, J.J. and Hou, Y.-M. (2008) Methods for kinetic and thermodynamic analysis of aminoacyl-tRNA synthetases. *Methods*, **44**, 100–118.
31. Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
32. Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978) In: *Statistics for Experimenters*. Wiley Interscience, NY.
33. Li, L., Francklyn, C. and Carter, C.W. Jr (2013) Aminoacylating urzymes challenge the RNA world hypothesis. *J. Biol. Chem.*, **288**, 26856–26863.
34. Jencks, W.P. (1987) In: *Catalysis in Chemistry and Enzymology*. Dover Publications, NY.
35. Horovitz, A. and Fersht, A.R. (1990) Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. *J. Mol. Biol.*, **214**, 613–617.
36. 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.
37. Tropsha, A., Carter, C.W. Jr, Cammer, S. and Vaisman, I.I. (2003) Simplicial neighborhood analysis of protein packing (SNAPP): a computational geometry approach to studying proteins. *Meth. Enz.*, **374**, 509–544.
38. Carter, C.W. Jr, LeFebvre, B.C., Cammer, S.A., Tropsha, A. and Edgell, M.H. (2001) Four-body potentials reveal protein-specific correlations to stability changes caused by hydrophobic core mutations. *J. Mol. Biol.*, **311**, 625–638.
39. Kapustina, M., Weinreb, V., Li, L., Kuhlman, B. and Carter, C.W. Jr (2007) A conformational transition state accompanies tryptophan activation by *B. stearothermophilus* tryptophanyl-tRNA synthetase. *Structure*, **15**, 1272–1284.
40. Sunden, F., Peck, A., Salzman, J., Ressler, S. and Herschlag, D. (2015) Extensive site-directed mutagenesis reveals interconnected functional units in the alkaline phosphatase active site. *Elife*, **4**, e06181.
41. Chandrasekaran, S.N., Das, J., Dokholyan, N.V. and Carter, C.W. Jr (2016) A modified PATH algorithm rapidly generates transition states comparable to those found by other well established algorithms. *Struct. Dyn.*, **3**, 012101.
42. Cammer, S. and Carter, C.W. Jr (2010) Six rosmannoid folds, including the class I aminoacyl-tRNA synthetases, share a partial core with the anticodon-binding domain of a class II aminoacyl-tRNA synthetase. *Bioinformatics*, **26**, 709–714.
43. Carter, C.W. Jr, Chandrasekaran, S.N., Weinreb, V., Li, L. and Williams, T. (2017) Combining multi-mutant and modular thermodynamic cycles to measure energetic coupling networks in enzyme catalysis. *Struct. Dyn.*, **4**, 032103.
44. Chandrasekaran, S.N. and Carter, C.W. Jr (2017) Adding torsional interaction terms to the Anisotropic Network Model improves the PATH performance, enabling detailed comparison with experimental rate data. *Struct. Dyn.*, **4**, 032103.
45. Weinreb, V. and Carter, C.W. Jr (2008) Mg<sup>2+</sup>-free *B. stearothermophilus* tryptophanyl-tRNA synthetase activates tryptophan with a major fraction of the overall rate enhancement. *J. Am. Chem. Soc.*, **130**, 1488–1494.
46. Benner, S.A., Sassi, S.O. and Gaucher, E.A. (2007) Molecular Paleoscience: systems biology from the past. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **75**, 9–140.
47. Benner, S.A. (1989) Enzyme kinetics and molecular evolution. *Chem. Rev.*, **89**, 789–806.
48. Liberles, D.A., Teichmann, S.A., Bahar, I., Bastolla, U., Bloom, J., Bornberg-Bauer, E., Colwell, L.J., Koning, A.P.J.d., Dokholyan, N.V., Echave, J. et al. (2012) The interface of protein structure, protein biophysics, and molecular evolution. *Prot. Sci.*, **21**, 769–785.
49. Bridgham, J.T., Ortlund, E.A. and Thornton, J.W. (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*, **461**, 515–519.
50. Ortlund, E.A., Bridgham, J.T., Redinbo, M.R. and Thornton, J.W. (2007) Crystal structure of an ancient protein: evolution by conformational epistasis. *Science*, **317**, 1544–1548.
51. Bridgham, J.T., Carroll, S.M. and Thornton, J.W. (2006) Evolution of hormone-receptor complexity by molecular exploitation. *Science*, **312**, 97–101.
52. Thornton, J.W. (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.*, **5**, 366–375.
53. Thornton, J.W., E., N. and D., C. (2003) Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science*, **301**, 714–717.
54. Martinez-Rodriguez, 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.
55. 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.
56. 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.
57. 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.
58. Carter, C.W. Jr and Wills, P.R. (2019) Experimental solutions to problems defining the origin of codon-directed protein synthesis. *Biosystems*, **183**, 103979.
59. Carter, C.W. Jr, Li, L., Weinreb, V., Collier, M., Gonzales-Rivera, K., Jimenez-Rodriguez, M., Erdogan, O. and Chandrasekharan, S.N. (2014) The Rodin-Ohno hypothesis that two enzyme superfamilies descended from one ancestral gene: an unlikely scenario for the origins of translation that will not be dismissed. *Biol. Direct*, **9**, 11.