

# Functional Class I and II Amino Acid-activating Enzymes Can Be Coded by Opposite Strands of the Same Gene<sup>\*[5]</sup>

Received for publication, February 2, 2015, and in revised form, June 1, 2015. Published, JBC Papers in Press, June 18, 2015, DOI 10.1074/jbc.M115.642876

Luis Martinez-Rodriguez, Ozgün Erdogan, Mariel Jimenez-Rodriguez, Katiria Gonzalez-Rivera<sup>1</sup>, Tishan Williams, Li Li, Violetta Weinreb, Martha Collier, Srinivas Niranj Chandrasekaran, Xavier Ambroggio<sup>2</sup>, Brian Kuhlman, and Charles W. Carter, Jr.<sup>3</sup>

From the Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7260

**Background:** Two distinct aminoacyl-tRNA synthetase classes may have descended from opposite strands of a single gene.

**Results:** Both products from a designed gene encoding ATP-binding sites from Class I and II synthetases catalyze amino acid activation.

**Conclusion:** The unique information in a gene can be interpreted two alternative ways, both of which accelerate similar chemistry.

**Significance:** Sense/antisense ancestry reshapes our assessment of the proteome.

Aminoacyl-tRNA synthetases (aaRS) catalyze both chemical steps that translate the universal genetic code. Rodin and Ohno offered an explanation for the existence of two aaRS classes, observing that codons for the most highly conserved Class I active-site residues are anticodons for corresponding Class II active-site residues. They proposed that the two classes arose simultaneously, by translation of opposite strands from the same gene. We have characterized wild-type 46-residue peptides containing ATP-binding sites of Class I and II synthetases and those coded by a gene designed by Rosetta to encode the corresponding peptides on opposite strands. Catalysis by WT and designed peptides is saturable, and the designed peptides are sensitive to active-site residue mutation. All have comparable apparent second-order rate constants  $2.9-7.0E-3 M^{-1} s^{-1}$  or  $\sim 750,000-1,300,000$  times the uncatalyzed rate. The activities of the two complementary peptides demonstrate that the unique information in a gene can have two functional interpretations, one from each complementary strand. The peptides contain phylogenetic signatures of longer, more sophisticated catalysts we call *Urzymes* and are short enough to bridge the gap between them and simpler uncoded peptides. Thus, they directly substantiate the sense/antisense coding ancestry of Class I and II aaRS. Furthermore, designed 46-mers achieve similar catalytic proficiency to wild-type 46-mers by significant increases in both  $k_{cat}$  and  $K_m$  values, supporting suggestions that the earliest peptide catalysts activated ATP for biosynthetic purposes.

Previous studies, summarized in Fig. 1 (1, 2), established a strong posterior probability of the sense/antisense ancestry hypothesis (3, 4) by verifying biochemical predictions of the hypothesis. Parallel structural evolution of the two families is consistent with a similar hierarchy over a  $10^5$ -fold range in catalytic proficiency. We deconstructed members of both aaRS<sup>4</sup> superfamilies to reveal structurally invariant cores that are relatively free of insertions and deletions that would rule out such coding (5, 6). These “*Urzymes*” (from *Ur* = primitive) contain the most highly conserved 120–130-amino acid fragments from the two superfamilies. They contain intact active sites, and they accelerate amino acid activation by  $10^8-10^9$ -fold (7, 8) and tRNA aminoacylation by  $10^6$ -fold (9). Furthermore, they are consistent with the antiparallel sequence alignment implied by the Rodin-Ohno hypothesis. This is not true for the intermediate sized catalytic domains with 200–325 residues, which exhibit somewhat larger catalytic enhancements (1, 10), but in Class I they contain long and variable insertions.

Bioinformatic evidence for sense/antisense ancestry is largely obscured by adaptive radiation of Class I and II genes to generate first a full repertoire of synthetases and then through the  $\sim 3.5$  billion years during which phyla speciated. We nonetheless identified statistically significant (34%) codon middle-base pairing between antiparallel sequence alignments excerpted from  $\sim 200$  contemporary Class I tryptophanyl-(TrpRS) and Class II histidyl (HisRS)-tRNA synthetases and smaller samples from the closely related tyrosyl- and prolyl-tRNA synthetase sequences. Those percentages increased to 42% as independently reconstructed ancestral sequences approached their respective roots (11).

\* This work was supported, in whole or in part, by National Institutes of Health Research Grant GM78227 from the NIGMS (to C. W. C. Jr.) and Grant GM073960 from the NIGMS (to B. K.). The authors declare that they have no conflicts of interest with the contents of this article.

[5] This article contains supplemental data.

<sup>1</sup> Present address: Dept. of Bacteriology, University of Wisconsin, Madison, WI 53706.

<sup>2</sup> Present address: Rosetta Design Group LLC, 47 Maple St. Suite 202, Burlington, VT 05401.

<sup>3</sup> To whom correspondence may be addressed. Dept. of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7206. Tel.: 919-966-3263; Fax: 919-966-2852; E-mail: carter@med.unc.edu.

<sup>4</sup> The abbreviations used are: aaRS, aminoacyl-tRNA synthetase; 46-mer, a general term describing specific 46-residue peptides corresponding to the ATP-binding sites of Class I and Class II aminoacyl-tRNA synthetases; TrpRS, tryptophanyl-tRNA synthetase; HisRS, histidyl-tRNA synthetase; Trp-46-mer, the specific ATP-binding site cloned from tryptophanyl-tRNA synthetase; His-46-mer, the specific ATP-binding site cloned from histidyl-tRNA synthetase; MBP, maltose-binding protein; PP<sub>i</sub>, pyrophosphate; PDB, Protein Data Bank; NTA, nitrilotriacetic acid.

Despite the strength of these two kinds of indirect evidence, it would be desirable to establish whether peptides translated from opposite strands of one gene can indeed accelerate the appropriate chemistry. We report such a direct experimental test here (Fig. 2), in which we extend the hierarchy in Fig. 1 to structures only 46 residues long.

Our earlier biochemical studies showed that Urzymes from both classes retain ~60% of the free energies of activation associated with  $k_{\text{cat}}/K_m$  (7–9) and ~25% of those associated with specific rejection by fully evolved synthetases of noncognate amino acids (1, 2). The catalytic power of peptides related by consensus phylogeny to contemporary aaRS is thus far greater than was anticipated from the gap between the uncatalyzed rates of amino acid activation and peptide bond formation. Uncatalyzed peptide bond formation (12, 13) is ~ $10^6$ -fold slower than the rate at which Urzymes produce activated amino acids. This million-fold excess catalytic proficiency argues that Urzymes closely resemble true ancestral forms, that they are highly evolved, and hence had simpler functional ancestors. We show here that such ancestors might now themselves be experimentally accessible, especially in context of the literature on uncatalyzed rates (12–16), which both motivated this study and helped to shape the methods used.

ATP binds strongly to specific locations in Class I and II aaRS (Fig. 2). The Class I-binding site (17) occurs in a 46-residue N-terminal fragment (46-mer) that has a catalytic signature containing the four amino acids, HIGH. The Class II-binding site contains an arginine-rich signature known as the motif 2 loop (18–21) near the C terminus of a comparable 46-mer. Both signatures have been implicated in catalysis of amino acid activation by full-length synthetases (7, 22–24). Moreover, they anchor sense/antisense alignments of corresponding 46-mers from the two aaRS classes, reducing ambiguity from insertions and deletions in designing a sense/antisense gene.

Earlier work by Mildvan and co-workers on P-loop peptides from  $F_1$ -ATPase (25, 26), DNA polymerase (27), and adenylate kinase (28, 29) showed that these three ~50-residue nucleotide-binding peptides exhibit ATP-dependent conformational transitions and have ATP affinities similar to those of the full-length enzymes.

Preliminary experiments, described at the beginning of “Results,” show that the wild-type 46-residue ATP-binding peptides cloned from *Geobacillus stearothermophilus* Class I TrpRS and *Escherichia coli* HisRS both have high ATP affinity and both accelerate amino acid-dependent  $^{32}\text{P}$ <sub>i</sub> exchange, encouraging us to carry out the protein design experiment reported here.

We adapted the Rosetta multistate protein design functionality (30–32) to constrain coding sequences for 46-residue Class I and II ATP-binding sites to be fully complementary. We describe here the isolation and functional characterization of 46-mers expressed from the designed sense/antisense gene, and we show that active-site mutations in both products eliminate catalytic activity of the designed peptides. Characterization of the wild-type 46-mers, together with the mutational inactivation and statistically significant dependences of the activities of the designed 46-mers on time, amino acid, and

peptide concentrations, compose substantive evidence for the authentic catalytic activities of both peptides.

## Materials and Methods

**Rates of Potential Competing Reactions**—We measured the  $^{32}\text{P}$ PP<sub>i</sub> exchange reaction catalyzed by 46-mer peptides on an ~14-day time scale, over which two competing reactions, uncatalyzed reaction of ATP with amino acid and hydrolysis of ATP itself, are potentially significant. We measured uncatalyzed amino acid-dependent synthesis of  $^{32}\text{P}$ ATP from unlabeled ATP and  $^{32}\text{P}$ PP<sub>i</sub> in five independent experiments. The five rates range over about an order of magnitude; averaging their free energies of activation yields a second order rate constant of  $4.2\text{E-}9\text{ s}^{-1}\text{ M}^{-1}$ , close to that ( $8.3\text{E-}9\text{ s}^{-1}\text{ M}^{-1}$ ) estimated (5) from model compound reactions by Kirby and Younas (14). ATP hydrolyzes spontaneously over the same time frame (16) with a reported half-life of 212 days ( $0.0033\text{ days}^{-1}$ ). We confirmed experimentally that a similar rate ( $0.0035\text{ days}^{-1}$ ) applied to our assay conditions.

**Computing Assay Blanks Reflecting the Uncatalyzed Rate**—Uncertainty of the background (*i.e.* the “blanks” subtracted from the raw counts/min) compounds the noise associated with enzymatic assays of amino acid activation by 46-mer peptides. We experimented with different ways to track the changing background over the ~2-week time course of the assay, measuring blanks separately at each time point in the assays of WT 46-mers.

For assays of the designed peptides we used the experimental determinations described above to compute time-dependent increase in disintegrations/min expected for uncatalyzed blanks as shown in Equation 1,

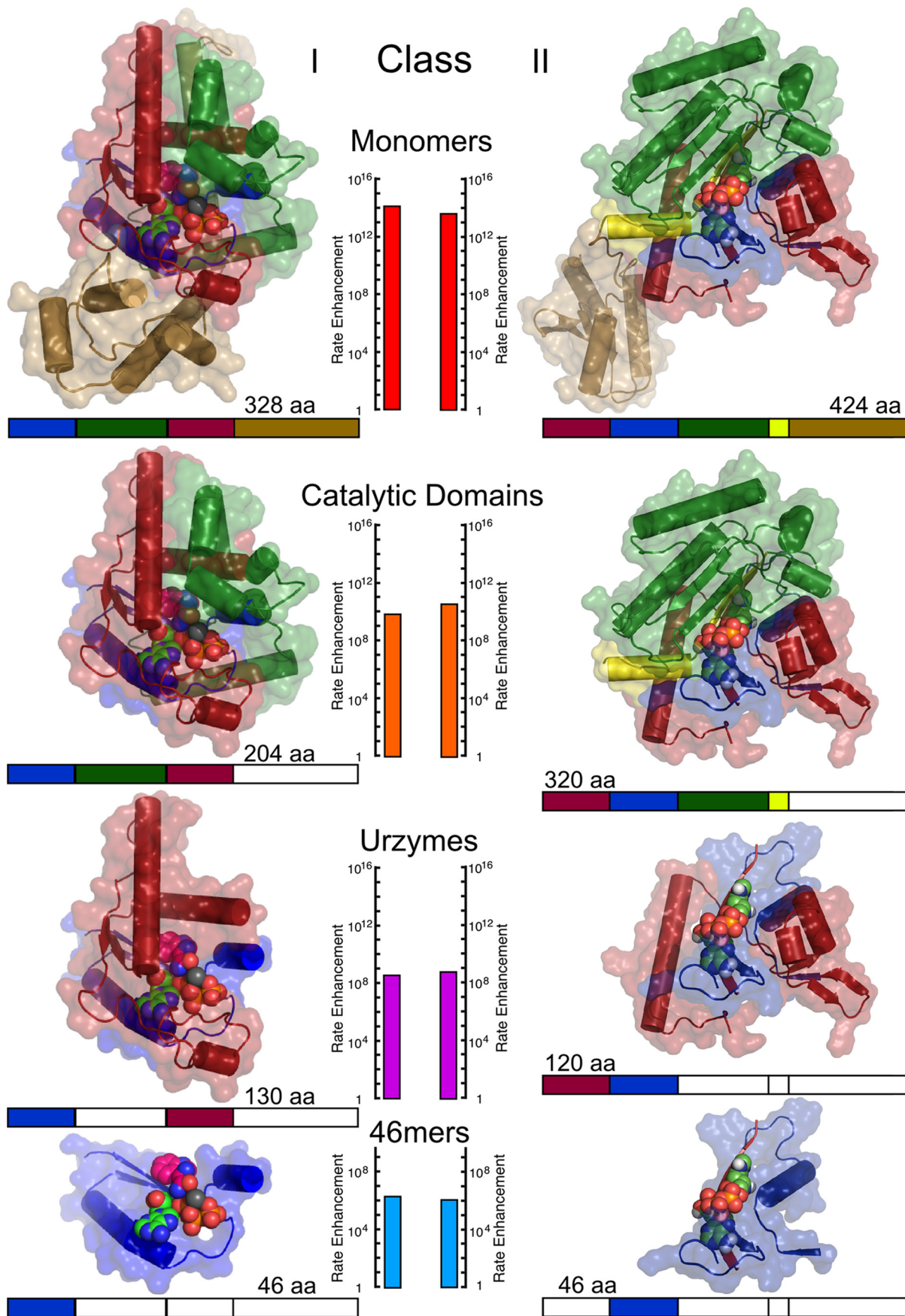
$$\text{blank}(t) = B_0 + \frac{(k_{\text{uncat}} \cdot t \cdot V \cdot [\text{aa}]) / \exp\left(-\frac{k_{\text{ATP}} \cdot t}{2}\right)}{4 \cdot C_0 / \exp(k_{32p} \cdot t)} \quad (\text{Eq. 1})$$

where  $B_0$  is estimated from the mean counts/min of blanks counted at time = 0;  $k_{\text{uncat}}$  is the uncatalyzed  $^{32}\text{P}$ PP<sub>i</sub> exchange rate constant;  $V$  is the reaction volume (200  $\mu\text{l}$ );  $k_{\text{ATP}}$  is the ATP hydrolysis rate constant;  $t$  is the time in seconds;  $k_{32p}$  is the  $^{32}\text{P}$  decay rate constant; aa is the amino acid concentration; and 4 reflects the 0.25 efficiency of ATP recovery from charcoal, determined separately.

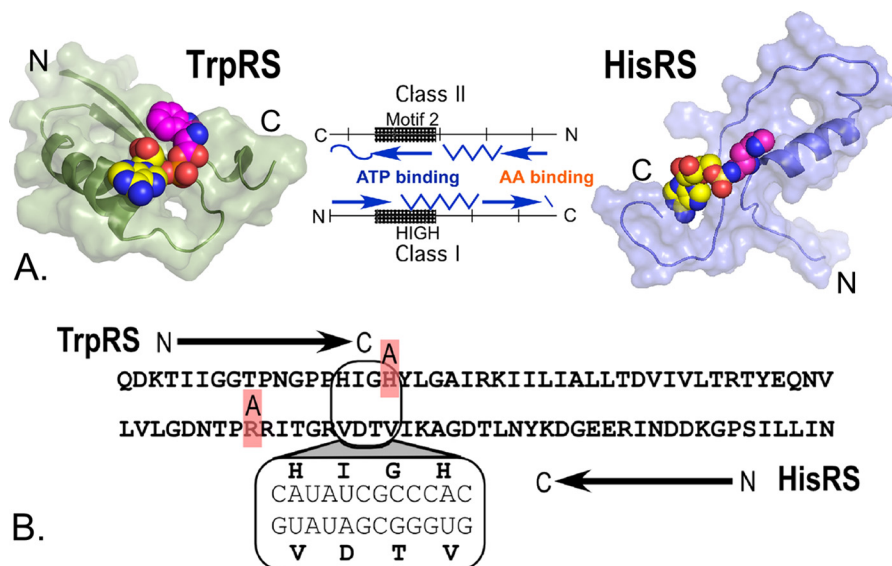
The  $^{32}\text{P}$ PP<sub>i</sub> exchange reaction starts with unlabeled ATP; the only label in ATP is thus that arising from exchange of labeled PP<sub>i</sub> into previously synthesized aminoacyl-5' AMP. The factor of 2 in the numerator of the exponential reflects the probability that  $^{32}\text{P}$ PP<sub>i</sub>, labeled in only one phosphate, will insert the label in the  $\gamma$ -position of ATP.  $C_0$ , the factor converting disintegrations/min to micromolar  $^{32}\text{P}$  at  $t = 0$  was determined by fitting standards counted at successive time points to a single exponential and extrapolating to zero time. This procedure reduced noise from the background.

**Design of the Sense/Antisense Gene**—The algorithm was essentially a parallel implementation of a multistate design as described (30, 33). Rosetta was configured to constrain the

Both Strands of a Gene Can Encode Functional Enzymes







**FIGURE 2. Sense/antisense gene coding for Class I and II ATP-binding sites from opposite strands.** *A*, schematic of gene architecture. ATP and amino acid-binding sites of the two antisense gene products reflect across the nucleic acid backbone symmetry axis (*center*), so that ATP-binding sites are at the N terminus of Class I and C terminus of Class II aaRSs. Structural schematics drawn from PDB files output by Rosetta for Class I (*left*) Class II (*right*) 46-mers to suggest possible folded configurations are in similar but not identical orientations to those in Fig. 1. These configurations reflect structural constraints input to Rosetta. Spheres are activated aminoacyl (magenta)-5'-AMP (yellow). *B*, sequences designed by Rosetta to stabilize tertiary structures in the *top* panel, preserving coding sequence complementarity. The *inset* shows the Class I catalytic HIGH and Class II motif 2 signatures; nucleic acid sequence of the gene is shown in Fig. 6*A*. Red letters denote mutations made to catalytic residues.

choice of amino acids at permitted positions on the opposite coding strand to those stabilizing backbones of the Class I and II ATP-binding sites derived from crystal structures of full-length TrpRS (1MAU, residues 1–46) and HisRS (1EL9, residues 77–122) bound to cognate aminoacyl-adenylate ligands. Consensus active-site residues Pro-10, His-15, Gly-17, and His-18 in the Class I and Arg-113 in the Class II peptide were already coded by complementary codons and were not allowed to vary.

Scoring according to the Rosetta energy function (34) was applied to amino acids for both backbone configurations. An external loop filtered selected amino acids to permit only those joint mutations that had complementary codons. The resulting gene sequence had the lowest Rosetta energy score from an ensemble of 50 designs. The selected gene (PDB files output by Rosetta are appended in the [supplement material](#)) was commissioned from GeneScript in two orientations, so that each peptide could be expressed and purified separately.

**Purification of Wild-Type 46-Residue Peptides**—We cloned, expressed, and purified the 46-residue ATP-binding peptides (*i.e.* not fusion proteins) from TrpRS and HisRS with His<sub>6</sub> tags. Peptides were solubilized from inclusion bodies in 6 M guanidinium hydrochloride and purified by affinity chromatography on nickel-NTA. They were dialyzed into 10 mM ammonium acetate and concentrated by lyophilization. Their short length makes it unlikely that they have unique conformations. We therefore assumed that denaturation did not alter their structural ensembles in buffer, and we assayed them without a renaturation step.

**Fluorescence Titration of ATP with WT Peptides**—ATP fluorescence quenching was measured as described for the 50-residue Walker A sequence of F<sub>1</sub>-ATPase (25), a putative homolog of the TrpRS 46-mer. Experiments were performed in 5 mM sodium acetate, pH 4.0, at 25 °C. TrpRS and HisRS 46-mer concentrations were increased from 0 to 510 μM and from 0 to 280 μM, respectively. Solutions were excited at 240 nm and recorded using a SPEX Fluorolog-3 spectrophotometer. The entire emission spectrum from 300 to 450 nm was recorded for buffer alone and for each peptide concentration in the presence and absence of 12.6 μM ATP (Trp-46-mer). ATP was increased to 120 μM for the His-46-mer to increase the signal.

Peptide/ATP interactions included both quenching of ATP by peptide and quenching of peptide fluorescence by ATP. The ATP concentration dependence of these changes was estimated by subtracting the integrated fluorescence of buffer alone from spectra with increasing concentrations of peptides. Difference curves were fitted to Equation 2,

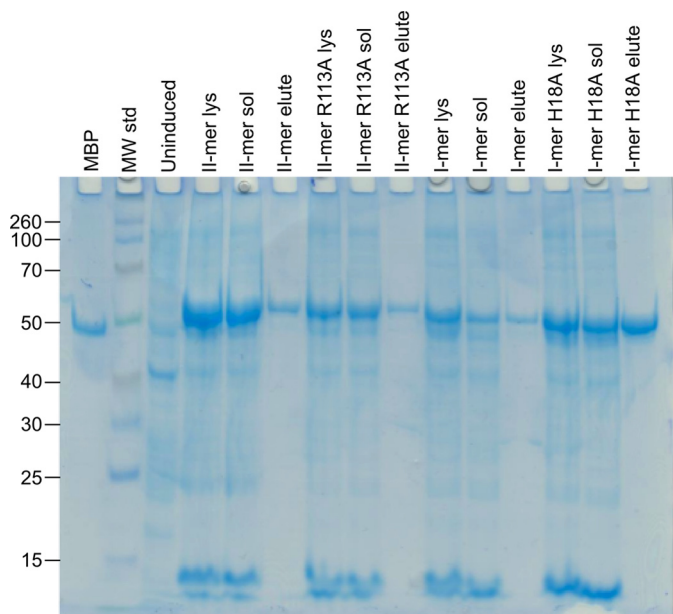
$$\Delta\Phi = 1 - \frac{A \cdot [46 \text{ mer}]}{[46 \text{ mer}] + K_{D, \text{ATP}}} \quad (\text{Eq. 2})$$

where  $A$  is an amplitude of ATP-dependent quenching, and  $K_{D, \text{ATP}}$  is the dissociation constant for the peptide-ATP complex.

**Expression of Designed Peptides**—The 46-mer peptides were expressed as MBP fusion proteins with a His<sub>6</sub> tag and purified from BL21 *E. coli* cells. Cells were transformed by heat shock

**FIGURE 1. Deconstruction of Class I tryptophanyl (PDB code 1MAU)- and II histidyl (PDB code 2EL9)-tRNA synthetases into successively smaller fragments that retain catalytic activity (1, 7–10, 43).** Graphics for smaller constructs are derived from coordinates of the full-length enzymes. Colored bars below each structure denote the modules contained within each structure; white segments are deleted. The number of amino acids (*aa*) in each construct is noted. Measured catalytic rate enhancements relative to the uncatalyzed second-order rate ( $k_{\text{cat}}/K_m/k_{\text{non}}$ ) are plotted on vertical scales aligned in the center of the figure and are colored from blue (slower) to red (faster). The 46-mers are the subject of this work.

## Both Strands of a Gene Can Encode Functional Enzymes



**FIGURE 3. Gel showing expression and purification of designed and active site mutants of the Class I (I-mer) and Class II (Il-mer).** MBP is maltose-binding protein; *Lys* is crude lysate; *sol* is soluble supernatant after lysis; and *elute* is the fraction eluted from nickel-NTA and assayed.

and left to recover for 1 h in LB growth media. They were grown overnight in media containing 100  $\mu\text{g/ml}$  ampicillin, transferred to 1 liter of fresh media, and incubated until the  $A_{600} = 0.6$ . Cells were harvested by centrifugation at 4,000 rpm at 4°C for 20 min and lysed in 50 mM  $\text{Na}_2\text{HPO}_4$ , pH 8.0, 300 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 0.5% v/v Nonidet P-40 by sonicating in three 45-s intervals with 15 s rest with cooling. Sonicated cells were centrifuged at 15,000 rpm for 20 min, and the supernatant was collected. Nickel-NTA beads were added to the supernatant and left overnight. The beads were washed once with lysis buffer and twice with 50 mM  $\text{Na}_2\text{HPO}_4$ , pH 7.0, 300 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 30 mM imidazole for 20 min each. Beads were transferred to a column, and protein was eluted using with 300 mM imidazole in five 2-ml fractions. Pooled fractions were dialyzed and stored at  $-20^\circ\text{C}$  in 50% glycerol and were assayed without tobacco etch virus protease cleavage. Concentrations of fusion peptide stocks were estimated from  $A_{280}$  and confirmed by standardized Coomassie staining on polyacrylamide gels (Fig. 3).

**Pyrophosphate Exchange by WT Peptides**—Pyrophosphate exchange, the standard assay for aminoacyl-adenylate (AA) synthesis, measures the incorporation of labeled inorganic pyrophosphate,  $^{32}\text{PP}_i$ , into ATP by reversal of the amino acid activation Reaction 1.



REACTION 1

The 2-week time courses equal the half-life of  $^{32}\text{P}$  and hence approach but do not exceed the detection limit.

$^{32}\text{PP}_i$  exchange assays were done at 37°C and initiated with 10  $\mu\text{l}$  of enzyme in 190  $\mu\text{l}$  of assay mixture: 0.1 M Tris-HCl, pH

8.0, 0.01 M KF, 5 mM  $\text{MgCl}_2$ , 1 mM ATP, and 10 mM tryptophan for Trp-46-mer or 10 mM histidine for His-46-mer, 70 mM 2-mercaptoethanol plus 2 mM  $^{32}\text{PP}_i$  at a specific radioactivity between 1 and  $2 \times 10^5$  cpm. Michaelis-Menten experiments were performed with 1 mM ATP and varying amino acid concentrations for amino acid-dependent kinetics, and 10 mM amino acid and varying concentrations of ATP for ATP-dependent kinetics. These conditions saturate the peptides in the fixed substrate.

All assays were processed by adding 200- $\mu\text{l}$  samples to 100  $\mu\text{l}$  of 7% perchloric acid on ice. A 15% activated charcoal suspension (50  $\mu\text{l}$ ) was added with mixing. The slurry was added to 0.2-mm centrifuge filter tubes and washed with 10 ml of water. Radiolabeled ATP was eluted with 100  $\mu\text{l}$  of pyridine and counted using Ecoscint A. Peptide concentrations of 100  $\mu\text{M}$  (Trp-46-mer) or 200  $\mu\text{M}$  (His-46-mer) and incubation times of 168 h were necessary for a sufficient signal-to-noise ratio from the WT peptides.

High tryptophan concentrations result in the elution from charcoal of a yellow scintillant. Thus, blanks increased linearly with tryptophan concentration. Blanks for the WT experiments were therefore prepared at different tryptophan concentrations and were used to construct a nomograph, from which appropriate blank values could be determined. To avoid this problem in assays of the designed peptides, leucine was used as the Class I amino acid.

**Pyrophosphate Exchange by Designed Peptides**—In view of the observation that very high concentrations of WT 46-mer were required in Michaelis-Menten experiments and that  $K_m$  values in those experiments were smaller for ATP than those for amino acids, we decided to carry out Michaelis-Menten experiments only for amino acid dependence of the designed 46-mers. Triplicate reactions for the designed peptides were carried out in sealed tubes with 1.0-ml volumes at  $\sim 35^\circ\text{C}$  in solutions of 50 mM Tris-HCl, pH 7.5, 10 mM KF, 2 mM ATP ( $>10 \times K_m$ ), 5 mM  $\text{MgCl}_2$ , 70 mM  $\beta$ -mercaptoethanol, 2 mM sodium pyrophosphate, to which were added  $^{32}\text{PP}_i$  to  $\sim 6 \times 10^5$  cpm/ml. Assays were done using two concentrations, 0.44 and 1.75  $\mu\text{M}$  of the respective peptide, and two nonzero concentrations, 10 and 32 mM leucine (Class I 46-mer) or histidine (Class II 46-mer). At appropriate time points, 200- $\mu\text{l}$  samples were withdrawn and treated as described for WT assays.

Designed peptides were assayed multiple times, including a single 8-day time point performed after the mutant peptides were assayed. A 16-day time point from that experiment exhibited similar WT sequence, peptide concentration, and amino acid concentration dependences but had slightly lower values than those from the 8-day time point and were discarded.

**Estimating Steady-state Parameters**—Catalytic proficiencies of WT and designed peptides can be compared with those of larger aaRS modules (Fig. 1) only by estimating their apparent second-order rate constants. The Michaelis-Menten formalism entails two unknown parameters that we wish to determine. The problem, especially with systems as noisy as those investigated here, is to estimate the two steady-state parameters with minimum uncertainty.

The 4-fold replication of assays for the WT peptides permitted the identification and elimination of  $\sim 10$  obvious outliers

before the remaining data were fitted to the Michaelis-Menten equation using nonlinear regression in JMP (35). High noise levels observed especially with ATP dependence in these assays suggested that a different strategy might be useful for assaying the designed peptides.

Aspects of this problem have been discussed, for example, by Hardin and Sloane (36) in the general context of statistical design and inference. A key strategy is to use high replication to increase the accuracy of experimentally determined points with the greatest influence on the model parameters.

As an example, consider the slope and intercept of an assumed linear relationship. Points close to the minimum and maximum of interest have the largest influence on the slope and intercept. Their values can be determined most accurately if experimental points at corresponding  $x$  values are replicated. Points in between have far less influence on either parameter than those at the extremes.

To estimate the parameters of a rectangular hyperbola, one wants to know as precisely as possible the reaction velocities at saturation (*i.e.* for  $k_{\text{cat}}$ ) and at half-saturation (*i.e.* for  $K_m$ ). Amino acid concentrations in our design (10 and 32 mM) approximate those values of the independent variable and amino acid concentration, and multiple measurements enhance the precision of measured velocities at those points by  $\sqrt{N}$  for  $N$  experimental determinations. Use of multiple regression methods, which allow all experiments to influence inferences about all factors, helps to realize this enhancement of precision in the estimation of coefficients. We therefore believe that this procedure gives a more accurate estimate of the steady-state parameters than would the conventional practice of fitting to measurements with fewer replications over a range of increasing substrate concentrations.

*Using Free Energies for Comparison*—We compare rate constants that span 14 orders of magnitude. The comparison benefits from converting both rate and binding constants to free energies. The free energy of activation contains an additive constant from the Boltzmann factor,  $\ln(kT/h = 6.2E^{12} \text{ s}^{-1})$ , which we omit, consistent with Ref. 37.

*Multiple Regression Methods*—Multivariate analysis is an extension of linear regression that makes the best use of appropriately balanced multivariate experimental designs (38, 39) necessary to characterize how phenomena depend on multiple independent variables.

$^{32}\text{P}i$  exchange activity may depend on the following independent variables: the class of the synthetase; whether the catalyst is a 46-mer or an MBP control; whether the sequences are designed (WT) or mutant; incubation time; peptide concentration; and the amino acid concentration. A fully balanced experimental design would test each combination of the above variables an equal number of times. We approximate that ideal in the experiment designed to test the activity of the designed 46-mers.

Three dependent variables were evaluated as follows (i) for  $^{32}\text{P}$ ATP, in micromolars, the amount of labeled ATP generated per experiment was used for regression modeling of the combined effects of all independent variables, and their time dependence gives estimates for rates of product formation for Michaelis-Menten analysis; (ii)  $^{32}\text{P}$ ATP produced per peptide

was used to assess distributions for WT, mutant, and MBP control experiments; and (iii)  $d[^{32}\text{P}]\text{ATP}/dt$ ,  $\text{s}^{-1}$  is the rate of  $^{32}\text{P}$ ATP synthesis for each data point. Each row of the experimental matrix associates these dependent variables for one observation of  $^{32}\text{P}i$  exchange activity with its own time in seconds, and numerical values indicating whether the peptide sequence was WT or mutant, whether it was an MBP control, peptide, and amino acid concentration, and class. These independent variables are denoted  $x_i$ .

Multivariate regression models (Equation 3) express calculated values for the experimental observations as linear combinations of the values of the independent variables:

$$\text{Prod}_{\text{calc}} = \beta_0 + \sum_{i=1,N}(\beta_i \cdot X_i) + \sum_{i=1,N;j=1,N}(\beta_{ij} \cdot X_i \cdot X_j) + \dots + \epsilon \quad (\text{Eq. 3})$$

where  $\text{Prod}_{\text{calc}}$  is the reaction product;  $\beta_0$  is the intercept where all  $x_i$  values are zero;  $\beta_i$  values are coefficients representing the slopes of the calculated function with respect to each independent variable,  $x_i$ ;  $\beta_{ij}$  are coefficients for interactions between  $x_i$  and  $x_j$ ; and  $\epsilon$  is a residual. A two-way interaction represents the different response to one variable at high and low values of another variable. Higher order interactions can be invoked if there is sufficient statistical support.

Coefficients are then estimated by least squares minimization of the difference between observed and calculated  $^{32}\text{P}$ ATP. The regression model itself is an equation for a multidimensional surface that best fits the observed data. If there are no two-way or higher interactions, this surface is a plane whose slopes in each direction are given by the  $\beta_i$ . Interaction terms introduce deformations that improve the fit of the surface to the data.

Regression modeling is a robust way to extract meaningful signals from noisy data. The advantages of this statistical procedure and a balanced design are as follows: first, each data point can be used together with all of the others to assess average differences at different values of the independent variables; and second, it affords a sound statistical test, the Student's  $t$  value and its probability under the null hypothesis, for the significance of each predictor,  $x_i$ ,  $x_i \cdot x_j$ , etc.

## Results

Absent a catalyst, amino acid activation is the slowest of the chemical steps necessary for protein synthesis and so is of central importance to the origin of biology. We describe catalysis of this key reaction by very short peptides. The experiments themselves and the conclusions that they support have little precedent. We highlight here the experimental observations supporting our conclusions, as well as the difficulties, the extent to which they have been overcome in the present experiments, and directions from which to seek further corroboration.

*Wild-type TrpRS and HisRS 46-mers Retain Functionality*—We first showed that peptides with wild-type sequences cloned by PCR from genes for the *Bacillus stearothermophilus* TrpRS and *E. coli* HisRS have ATP binding properties comparable with those found by Mildvan and co-workers (28–31) for peptides homologous to the Class I ATP-binding sites from several



## Both Strands of a Gene Can Encode Functional Enzymes

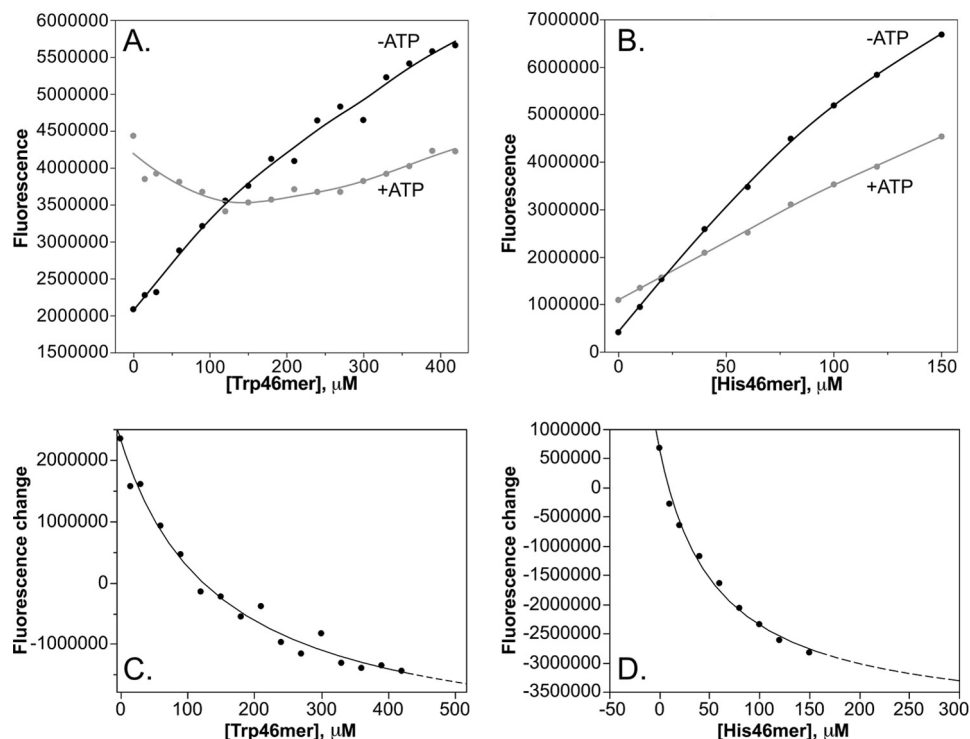


FIGURE 4. Fluorescence changes due to the presence of ATP, derived from titrations of ATP with TrpRS (A and C) and HisRS (B and D) 46-mer peptides. Integrated (300–450 nm) spectra for peptides without ATP (A and B, black) were subtracted from those with ATP (A and B, gray) to estimate quenching of ATP fluorescence by the peptides and of the peptides by ATP. ATP-dependent fluorescence changes were fitted to Equation 1 for Trp-46-mer (C) and His-46-mer (D). Dashed lines indicate data points omitted from fitting because of development of turbidity developing at high [peptide]. Fitted dissociation constants are given in the text.

ATP-dependent enzymes, and we then documented their catalytic activities.

$F_1$ -ATPase, adenylate kinase, and DNA polymerase I have P-loops and are distant homologs of the Class I ATP-binding site. Those peptides vary in length from 45 to 51 residues and were shown to bind ATP tightly ( $\sim 10 \mu\text{M}$ ) by titrating ATP with peptide and to undergo ATP-dependent conformational transitions by two-dimensional NMR (25–29).

The wild-type sequences for the TrpRS and HisRS 46-mers bind ATP tightly (Fig. 4). Fluorescence changes due to ATP exceed the total fluorescence from ATP alone, so they include both quenching of ATP fluorescence by Trp-46-mer and His-46-mer and quenching of peptide fluorescence by ATP. We subtracted titrations of buffer with peptide from titrations of ATP with peptide (Fig. 4, A and B) to identify these ATP and peptide concentration effects (Fig. 4, C and D).

Gel filtration experiments (data not shown) showed ATP-dependent oligomerization when either peptide was at high concentrations. These experiments dilute the peptides by more than an order of magnitude, and therefore assaying the monomer and dimer fractions was unreliable. Data points at high peptide concentrations were therefore excluded from fitting (dashed line segments in Fig. 4, C and D).

ATP-dependent fluorescence changes were fitted to Equation 1, assuming one site (Fig. 4, C and D), to give dissociation constants of  $145 \pm 19 \mu\text{M}$  for the Trp-46-mer and  $56 \pm 8 \mu\text{M}$  for the His-46-mer. These dissociation constants are severalfold lower than ATP-dependent  $K_m$  values for the full-length enzymes ( $300 \mu\text{M}$  for full-length TrpRS (9) and  $890 \mu\text{M}$  for full-length HisRS (40)), which

also are thermodynamic dissociation constants by virtue of the exchange condition of the assay (Cleland *et al.* (56)). High ATP affinity was observed by Mildvan and co-workers (25) for the corresponding ATP synthase Walker A peptide.

**Steady-state Kinetics Analysis of WT 46-mers**—The WT 46-mers also accelerate cognate amino acid activation. We did 168-h time course assays of amino acid-dependent  $^{32}\text{P}_i$  exchange for both WT 46-mers. Catalytic activity was significant by these tests; regression of 18 time-dependent measurements with and without peptide showed that Student's *t* test *p* values were  $\sim 10^{-4}$  for peptide catalysis and 0.01 for the effect of time in both cases.

Substrate-dependent assays (Fig. 5) show that both WT 46-mers exhibit saturation behavior with respect to both amino acid and ATP concentrations. The high peptide concentrations required for sufficient signal-to-noise in these measurements (shown by **bold arrows** in Fig. 5) place upper bounds on the Michaelis constants derived from fitting by nonlinear regression to rectangular hyperbolae. The apparent second-order rate constants for amino acids,  $k_{\text{cat}}/K_m = 3.5\text{E-}3 \text{ s}^{-1} \text{ M}^{-1}$  for Trp-46-mer and  $8.3\text{E-}4 \text{ s}^{-1} \text{ M}^{-1}$  for His-46-mer, are therefore lower limits for 46-mer catalytic proficiency. These limitations do not compromise the conclusion that although turnover is weak (average  $k_{\text{cat}} \sim 2.7 \pm 2/8\text{E-}7 \text{ s}^{-1}$ ; Table 1), the 46-mers are at least  $10^{-3}$ – $10^{-4}$  as proficient as the corresponding 120–130 residue Urzymes (9).

Michaelis constants for ATP are significantly smaller than those of the wild-type enzymes, confirming conclusions drawn from ATP fluorescence in peptide titrations. Moreover, those binding

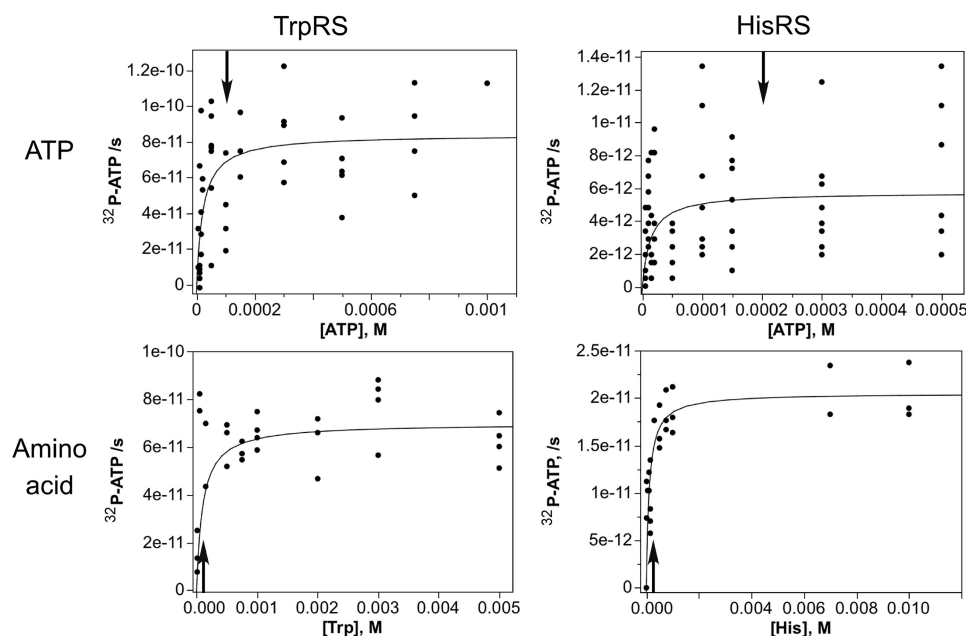


FIGURE 5. **Steady-state kinetic parameters for the WT TrpRS and HisRS 46-mer peptides.** Assays were conducted as described under “Materials and Methods.” Data points to the left of vertical lines at 100 mM along the x axis cannot be interpreted unambiguously, because the substrate concentrations are below the catalytic concentration. Velocities are therefore underestimates.

**TABLE 1**

**Steady-state kinetic parameters for wild-type TrpRS and HisRS 46-mers, derived from the fits shown in Fig. 5**

Standard errors for estimated parameters  $k_{cat}$  and  $K_m$  are calculated by the nonlinear least squares algorithm in JMP (35). The values for  $k_{cat}/K_m$  are derived from those standard errors by conventional propagation of errors for a ratio as follows:  $Std\_error(k_{cat}/K_m) \approx (k_{cat}/K_m) \cdot \sqrt{((j k_{cat}/k_{cat})^2 + (K_m/K_m)^2)}$ .

	Parameter	Estimate	Est Std Err
TrpRS		Tryptophan	
	$k_{cat}$ , $s^{-1}$	3.5E-07	3.3E-08
	$K_m$ , M	1.0E-04	5.9E-05
	$k_{cat}/K_m$ , $s^{-1}M^{-1}$	3.5E-03	2.1E-03
TrpRS		ATP	
	$k_{cat}$ , $s^{-1}$	5.9E-07	3.2E-08
	$K_m$ , M	2.9E-05	7.9E-06
	$k_{cat}/K_m$ , $s^{-1}M^{-1}$	2.0E-02	9.5E-03
HisRS		Histidine	
	$k_{cat}$ , $s^{-1}$	9.4E-08	9.4E-09
	$K_m$ , M	1.1E-04	5.1E-05
	$k_{cat}/K_m$ , $s^{-1}M^{-1}$	8.3E-04	3.8E-04
HisRS		ATP	
	$k_{cat}$ , $s^{-1}$	5.1E-08	5.6E-09
	$K_m$ , M	3.5E-05	1.3E-05
	$k_{cat}/K_m$ , $s^{-1}M^{-1}$	1.5E-03	5.6E-04

isotherms (Fig. 4) cannot be attributed to a minor contaminant. This, together with the fact that the peptides were isolated without renaturation from inclusion bodies, and therefore should not contain active wild-type enzymes, argues that the catalysis is not due to tiny amounts of contaminating wild-type activity.

**Designed Gene Encoding Class I and II 46-mers from Opposite Strands**—It is a straightforward extension of multistate design (30, 33) to build sequences simultaneously to stabilize two different backbone scaffolds. We configured this feature of the protein design program Rosetta to enforce complementary coding sequences consistent simultaneously with structures of the Class I and II ATP-binding 46-mers (Figs. 2 and 6). Roset-

ta's semi-empirical energy function (30, 31) makes its choices an effective surrogate for sequence information inaccessible from multiple sequence alignments of related proteins (Fig. 6B). The selected gene sequence, Fig. 6A, was synthesized (GeneScript) and cloned in opposite directions to express each peptide as a maltose-binding protein fusion with an additional His<sub>6</sub> tag.

**Class I and II Sense/Antisense MBP-46-mer Fusions Both Accelerate Amino Acid Activation**—Amino acid activation by designed and mutant peptides was followed by the  $^{32}P$ PP<sub>i</sub> exchange assay at different time points at different peptide and amino acid concentrations. This experimental matrix was used to assess the significance of answers to the following questions concerning the catalytic activity of the two peptides (Table 2). (i) Do levels of [ $^{32}P$ ]ATP increase significantly with time? (ii) Are rate accelerations observed for the two peptides significantly greater than those observed in the presence of MBP controls? (iii) Are the rate accelerations sensitive to active-site mutations? (iv) Do the rates depend on the concentrations of peptide and of substrate?

We made alanine substitutions to the second histidine of the HIGH signature of the Class I 46-mer (H18A) and arginine (R37A) of the Class II 46-mer (Fig. 2B) to authenticate these catalytic activities. The designed peptides were assayed initially and were re-purified and assayed again later, after the mutant peptides were assayed.

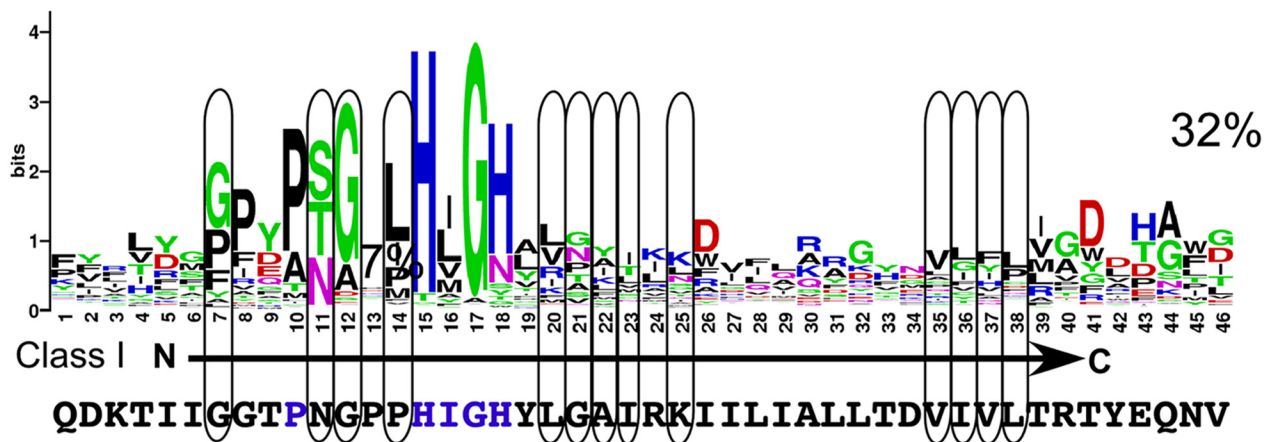
As the experimental data approach the level of detection, designed Class I and II 46-mers were assayed 96 times (six replicates, four times, two peptide concentrations, two amino acid concentrations), and the corresponding mutant peptides were assayed 48 times. This high replication increases the signal-to-noise ratio by  $\sqrt{N} \sim 10$ - or 7-fold for the mutants, affording a more precise determination of the uncertainty in average values estimated over multiple measurements.



## Both Strands of a Gene Can Encode Functional Enzymes

TrpRS 5' Q D K T I I G G T P N G P P H I G H Y L G A I  
 HisRS 3' CAGGACAAGACCATCATTGGTGGGACGCCCTAATGGTCCCCGCACATCGGTCACTATCTTGGCGCCATC  
 HisRS 3' GTCCTGTTCTGGTAGTAACCACCTGCGGATTACCAGGGGCGTGTAGCCAGTGATAGAACC GCGGTAG  
 L V L G D N T P R R I T G R V T D V I K A G D  
 R K I I L I A L L T D V I V L T R T Y E Q N V  
 CGTAAGATTATACTTATCGCCCTCCTCACGGATGTTATCGTCTTACCAGGACTTATGAGCAGAATGTT TrpRS 3'  
 GCATTCTAAATAGAATAGCGGGAGGAGTGCCTACAATAGCAGGAATGGTCTGAATACTCGTCTTACAA HisRS 5'  
 T L N Y K D G E E R I N D D K G P S I L L I N

A.



B.

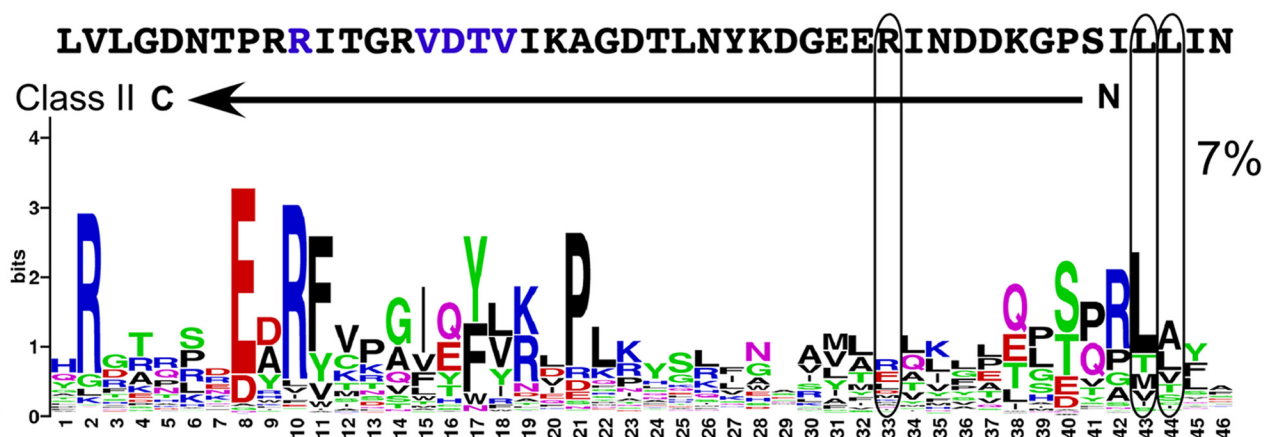


FIGURE 6. **Sense/antisense gene.** *A*, DNA sequence, with codon/anticodon pairs translated. *B*, comparison of the sense antisense gene designed by Rosetta (30) (center) with Logos derived from multiple alignments for ~200 sequences for each of six contemporary Class IA, -B, and -C and five Class IIA, -B, and -C sequences (Fig. prepared using Canvas 8.0 and WebLogo (55)). Rosetta was not allowed to vary the five blue amino acids in the designed sequence. Ovals are positions where, elsewhere, Rosetta chose consensus amino acids from the biological MSA. Percentages to the right indicate a wide variance in such selection between the Class I and II 46-mers as discussed in the text.

TABLE 2

Multivariate regression analysis of factors affecting amino acid activation by designed Class I and Class II 46-residue peptides containing ATP-binding sites

$R^2 = 0.71$  with 227 degrees of freedom;  $F = 61.0$  for all 237 observations from two investigators on different occasions ( $P(F) < 0.0001$ ). Coefficients are sorted in descending order of statistical significance given by  $p$  value, excluding the constant term.  $p$  values given are those for the  $F$  ratio statistic from stepwise multiple regression to emphasize the relative importance of different coefficients. These are approximately equal to the  $p$  values of the Student's  $t$  test.

Term	Estimate	Standard error	$t$ ratio	Probability $> F $
WT	7.7E-03	5.4E-04	14.25	2.20E-45
[Peptide], M	4.5E+03	4.2E+02	10.66	4.50E-35
(Time, s)·(WT)	1.5E-08	1.6E-09	9.78	1.70E-24
([Peptide], M)·(WT)	7.4E+03	8.0E+02	9.32	4.50E-19
Time, s	7.1E-09	7.8E-10	9.09	1.10E-17
([Peptide], M)·(class)	-5.9E+03	7.9E+02	-7.5	3.80E-15
MBP	-8.8E-03	1.4E-03	-6.43	1.50E-12
Class	-2.8E-03	5.2E-04	-5.42	7.40E-10
[Amino acid], M	4.2E-02	2.2E-02	1.94	5.39E-02

The redundancy and balance of the experimental design allowed us to verify that all dependences also are statistically significant (Table 2). Multiple regression methods average over

the high noise level, to assess the relative statistical significance simultaneously for differences in the time-dependent [ $^{32}$ P]ATP synthesis that answer questions i–iv. To assess the joint statis-

tical significance of our conclusions, columns of the design matrix were used to build regression models (see Equation 3) for the data measured by both investigators for WT and mutant peptides and MBP controls. All calculations were done using JMP (35).

Results of regression modeling in Table 2 convey in compact form many related conclusions that are not immediately apparent but that have bearing on the questions addressed by this work. The  $R^2$  value, 0.71, tells us that the model in Equation 3 can explain more than 70% of the variation in observed [ $^{32}$ P]ATP values. The unexplained variance approaches the noise level. As the ratio of a coefficient to its standard error, the Student's  $t$  values measure the signal-to-noise of the regression coefficients. Their  $p$  values are the probability of a Student's  $t$  value as large as that of the model coefficient under the null hypothesis that the different experimental data do not arise from differences in the corresponding independent variable. The relative impact of the noise depends on the magnitude of the experimental signal. Small  $p$  values mean that, despite the high noise level, the range of measurements is sufficient to exhibit the dependences we hoped to demonstrate. The  $p$  values thus furnish quantitative estimates of the impact of noise and certify the significance of differences despite the noise in the data.

All main (*i.e.* intrinsic) effects have  $p > |t|$  values  $\ll 0.0001$ . Their signs and magnitudes are physically reasonable. The most significant predictors of activity in Table 2 are the wild-type sequence (WT:  $t = 14$ ;  $p > |F| < 10^{-45}$ ), the peptide concentration ( $[\text{peptide}]$ ,  $M$ ;  $t = 10.7$ ;  $p > |F| < 10^{-35}$ ), the difference between the time dependence of the WT and mutant sequences ( $(\text{time}, s) \cdot (\text{WT})$ ;  $t = 9.8$ ;  $p > |F| < 10^{-24}$ ) and the time (time,  $s$ ;  $t = 9$ ;  $p > |F| < 10^{-17}$ ). All but one coefficient have  $p > |t|$  values  $\ll 0.0001$ .

Figs. 7 and 8 unpack the two most important conclusions from Table 2 graphically. Fig. 7 shows that (i) designed peptides do accelerate amino acid activation whereas MBP, expressed and purified by the same procedure as that used for the peptides, does not; and (ii) active-site mutations substantially reduce catalysis, thereby implicating the peptides themselves as the authentic catalysts.

Whereas the designed peptides are the dominant proteinaeous components of the solutions we assayed, there are doubtless other components, to which the observed rate accelerations might be attributed. These include possible contamination of the peptides by tiny amounts of wild-type aaRS not excluded during purification. As these concentrations also increase with the volume of catalyst added ( $[\text{peptide}]$ ), the statistical significance of that predictor alone does not, by itself, address the question of contamination.

Fig. 8 shows the increased rates of [ $^{32}$ P]ATP synthesis by different concentrations of the two designed peptides at the amino acid concentration = 0.032 M. The increased velocities at the higher peptide concentrations are also evidenced by the mean velocities (Table 3), which have been normalized for peptide concentration and are within experimental error. Increased peptide concentrations do not increase velocities either for the mutated peptides or for MBP only controls. That observation, which is confirmed quantitatively by the signifi-

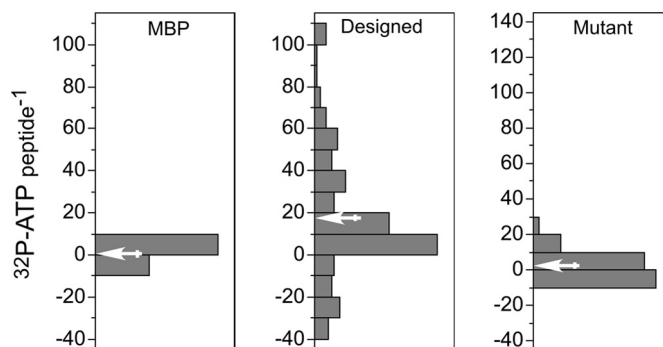


FIGURE 7. Distributions of [ $^{32}$ P]ATP produced per peptide for all assays for the designed 46-residue sense/antisense Class I and II peptide catalysts, MBP controls, and mutant peptides. Distributions of [ $^{32}$ P]ATP produced per peptide for all assays for the designed 46-residue sense/antisense Class I and II peptide catalysts, MBP controls, and mutant peptides are shown. All data were divided by concentration. White arrows denote averages of the distributions, and vertical cross-bars at the tails of the arrows give their standard errors of the mean. The y axis corresponds to the  $^{32}\text{PP}_i$  exchange data, in units of mole of ATP/mol of peptide. Left, MBP controls. Center, distributions for designed peptides. Right, mutant peptides. Differences between peptide and MBP controls have Z-scores of 10.4 and 9.8. The differences for Class I and II peptides have Z-scores of 3.0 and 11.4. Prepared using JMP (38).

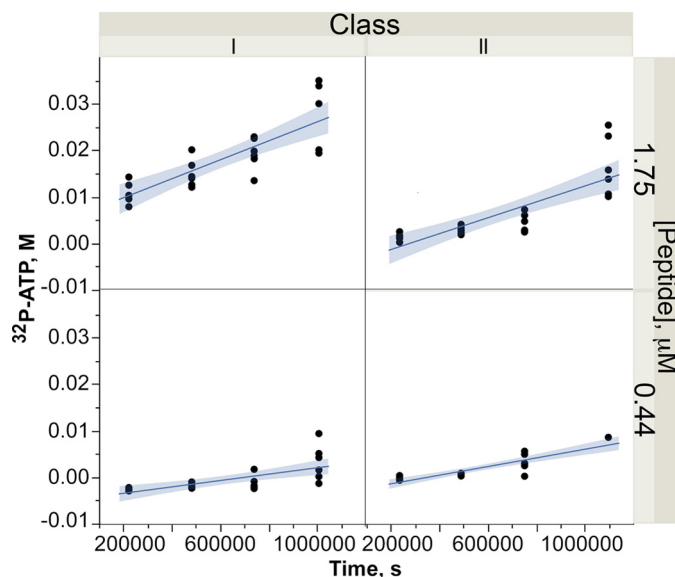


FIGURE 8. Time, class, and peptide concentration (0.44  $\mu\text{M}$ ; 1.75  $\mu\text{M}$ ) dependences of catalytic peptide activity. The y axis is in [ $^{32}$ P]ATP concentration. Variances of the time dependences are indicated by shading along either side of the fitted lines. Prepared using JMP (35).

cance of the  $[\text{peptide}] \cdot \text{WT}$  coefficient in Table 2, is evidence that the observed rate accelerations in Fig. 8 do not arise from contaminating activities, which would also increase activities in those controls. Table 2 thus simultaneously documents that solutions containing the designed peptides accelerate amino acid activation by different amounts in proportion to time, peptide, and amino acid concentrations and, especially, only with WT active-site residues.

The time dependences of the distinct MBP and mutant controls are both slightly negative ( $-6 \pm 2.2 \text{ E-}11 \text{ s}^{-1}$ ). However, their absolute values are  $\sim 2\%$  of that from WT sequences ( $3.5 \text{ E-}9 \pm 1.2 \text{ E-}9 \text{ s}^{-1}$ ) and hence are essentially independent of time. The negative time dependences are quite similar, and we conclude that our estimate of the uncatalyzed rate is a good

## Both Strands of a Gene Can Encode Functional Enzymes

**TABLE 3**

**Mean values and standard errors of rate data used to evaluate steady-state kinetic parameters for the designed peptides**

Average rates were estimated by fitting the time-dependent values for each [amino acid] and each [peptide] for each class over individual data points after correcting for peptide concentration and are plotted in Figs. 8, A–D. S.E. = standard error;  $n$  = number of replicates.

[Amino acid]	$n$	$^{32}\text{P}$ ATP_Leu	S.E._Leu	$^{32}\text{P}$ ATP_His	S.E._His
$M$		$s^{-1}$		$s^{-1}$	
0	36	8.5E-07	1.3E-06	3.4E-07	9.5E-07
0.44 $\mu\text{M}$ Peptide					
0.01	36	3.4E-05	2.0E-05	2.0E-05	6.4E-06
0.032	36	5.0E-05	1.4E-5	2.8E-05	5.9E-06
1.75 $\mu\text{M}$ Peptide					
0.01	36	4.6E-05	6.5E-06	0.00001	6.0E-06
0.032	36	7.4E-05	1.2E-05	2.6E-05	4.0E-06

approximation. For these reasons, the data in Table 2 argue significantly in favor of unique catalysis by the peptides themselves.

**Estimation of Apparent Second-order Rate Constants,  $k_{cat}/K_m$** —The Michaelis-Menten plots in Fig. 5 suggest that the WT 46-mer peptides saturate at high substrate concentrations. It is unlikely that such simple peptides can exhibit cooperative behavior with respect to substrate concentrations. Cooperative catalytic behavior can be difficult to establish even for full-length enzymes. Thus, Michaelis-Menten behavior is the simplest postulated behavior, and we assume it here for the purpose of comparing proficiencies with those of other previously characterized aaRS constructs (7–9).

Activities for both designed peptides also appear to saturate. In view of the intrinsically noisier plots of the [ATP] dependence of WT activities, we decided to fit to more highly replicated rates at three amino acid concentrations (0, 10, and 32 mM), shown with standard errors of the means in Table 3, rather than dividing them between amino acid and ATP experiments.

We estimated the steady-state parameters from the full data set by calculating specific velocities,  $d[^{32}\text{P-ATP}]/dt s^{-1}$ , for both peptide concentrations, and fitting to the Michaelis-Menten equation to obtain minimum variance estimates for the apparent second-order rate constants (Fig. 9). Data for each peptide concentration constitute independent experiments; agreement between these four experiments, each of two concentrations of the Class I (Fig. 9, A and C) and II (Fig. 9, B and D) peptides, provides estimates of consistency for each peptide, which is expected, and between estimates for the two Classes, which is unexpected.

The fits are in reasonably good agreement (Table 4). Michaelis-Menten parameters afford average estimates,  $7.7 \pm 0.9E-3 s^{-1} M^{-1}$  for Class I and  $3.3 \pm 1E-3 s^{-1} M^{-1}$  for Class II apparent second-order rate constants,  $k_{cat}/K_m$ , for amino acid activation. Thus, the biochemical behavior of the Class I and II 46-residue peptides encoded by opposite strands of the designed gene (Fig. 2B) exhibits all appropriate characteristics of enzymes; activities of both peptides are greatly reduced in active-site mutants, and  $K_m$  values for amino acids are significantly different from those of full-length enzymes. We conclude that the observed activities represent the authentic catalytic activities of gene products coded by opposite strands of the same gene.

**Design Altered the Enzymatic Properties of the Sense-Antisense Class I and II 46-mers**—WT and designed peptides have approximately the same catalytic proficiencies, as given by their activation free energies for  $k_{cat}/K_m$  ( $\sim 3.5$  kcal/mol; Table 5). Qualitatively, however, the WT peptides required  $\sim 100$ -fold higher peptide concentrations than the designed peptide assays to achieve adequate signal-to-noise ratios. That qualitative discrepancy implies significant differences in individual Michaelis-Menten parameters.

Free energies estimated from the steady-state parameters from six different determinations (TrpRS and HisRS WT 46-mers with ATP and amino acid, plus the Class I and II designed 46-mers with amino acid; Table 5) cluster significantly according to WT *versus* designed classes and substrates (Fig. 10). The two WT 46-mers bind substrates  $\sim 100$ -fold more tightly but have  $\sim 100$ -fold smaller turnover numbers than do the designed Class I and II 46-mers (Tables 1 and 4).

Because these differences are so large, their statistical significance is high (Tables 5 and 6). The design process consistently increased turnover number and reduced amino acid affinity. It remains to be tested whether or not the striking differences observed here result from imposing the sense/antisense constraint. An interesting implication is that changes required for sense/antisense coding may be inconsistent with high amino acid specificity.

This unexpected result is, however, consistent with the possibility that sense/antisense-encoded 46-mers functioned initially to bind ATP and to facilitate its reaction with low molecular weight substrates during the origin of life and that their specialization as amino acid-activating enzymes developed later (41). The surprisingly balanced catalytic behavior of the products of the designed gene also indirectly validates the novel use of Rosetta in the manner described here. The designed 46-mers behave much more like generalized ATP activators. The smaller turnover numbers of the WT 46-mers, in contrast, likely reflect the fact that much of the contemporary aaRS catalytic machinery involves interactions between the ATP-binding sites and other modular components (42, 43), which have been deleted.

The explanation that differences in the active fraction of molecules purified from inclusion bodies (wild type) and as fusion proteins (designed) account for these kinetic differences is rendered unlikely by the ATP quenching behavior of the wild-type peptides, which are inconsistent with activity vested in 1% of active molecules, as that would imply 100-fold tighter dissociation constants than those in Table 1.

**Comparison of Biological and Rosetta Multiple Sense/Antisense Sequence Alignments with the Assayed Sense/Antisense Sequences**—Many layers of amino acid substitutions were probably necessary to embed the putative primordial 46-residue catalysts represented here by the designed sequences into the contexts of the 10 contemporary aminoacyl-tRNA sequences that we propose descended from each strand. The gene sequence in Fig. 6A resulted from an initial design involving 50 trials. The resulting sequence is compared in Fig. 6B to sequences from Fig. 1 and to those from a multiple alignment of  $\sim 200$  biological sequences from six (LeuRS, IleRS, GluRS, GlnRS, TrpRS, and



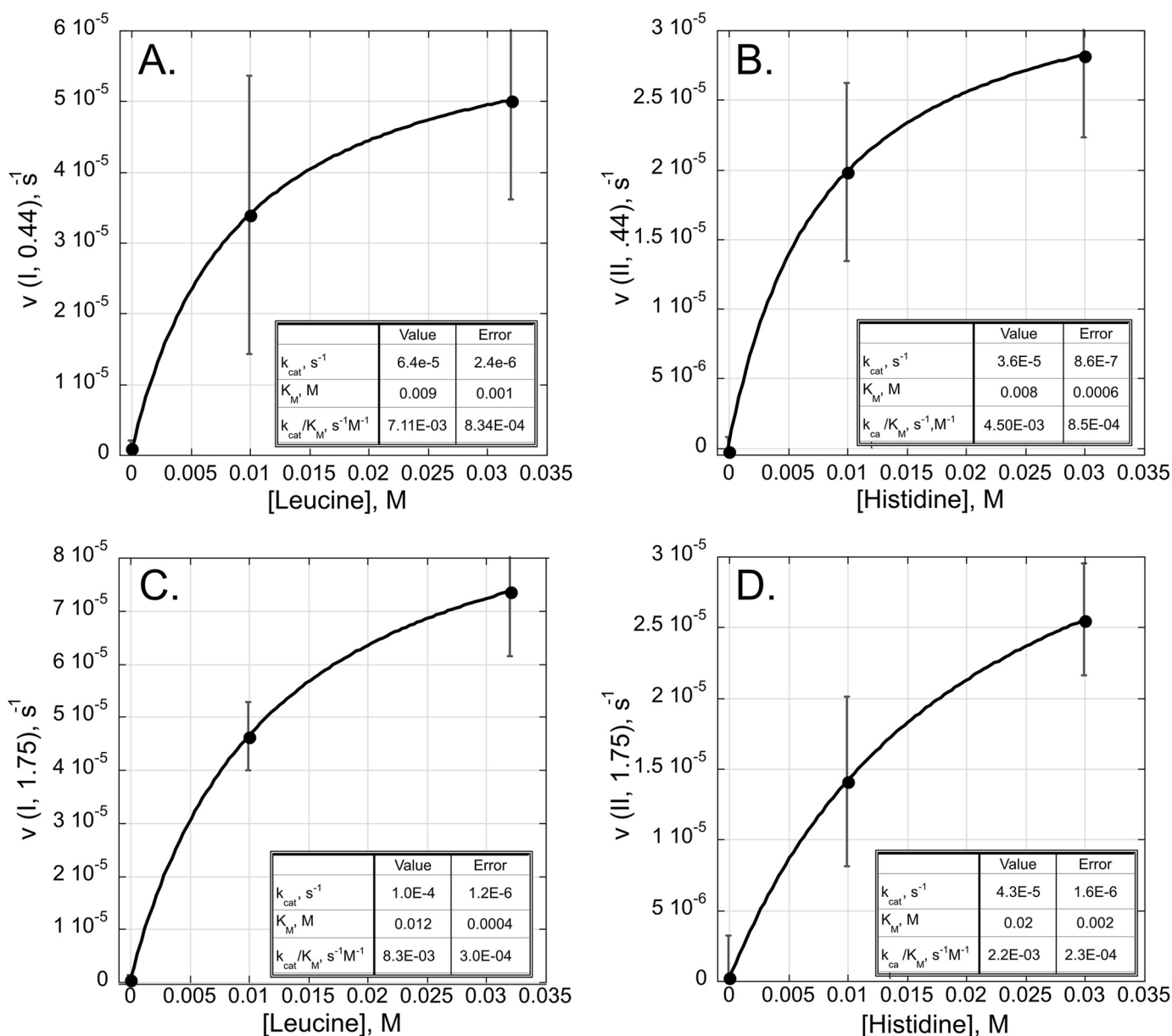


FIGURE 9. Estimation of apparent second-order rate constants. Amino acid concentration dependences plotted in Michaelis-Menten format. A–D, data for all time points for each unique combination of peptide and substrate concentrations were fitted against time in seconds by linear least squares. The mean values of rates for both amino acid concentrations were plotted against amino acid concentration for experiments with 0.44  $\mu\text{M}$  (A and B) and 1.75  $\mu\text{M}$  peptide (C and D), giving two independent experimental estimates of steady-state kinetic parameters for Class I (A and C) and Class II (B and D) peptides, which are given in Table 4.

TABLE 4

Apparent second-order rate constants,  $k_{\text{cat}}/K_m$  for Class I and II designed 46-mers

Standard errors (S.E.) for estimated parameters  $k_{\text{cat}}$  and  $K_m$  are calculated by the nonlinear least squares algorithm in JMP (35). Those for  $k_{\text{cat}}/K_m$  are derived from those standard errors by conventional propagation of errors for a ratio:  $\text{Std\_error}(k_{\text{cat}}/K_m) \sim (k_{\text{cat}}/K_m) \cdot \sqrt{(\sigma_{k_{\text{cat}}}/k_{\text{cat}})^2 + (\sigma_{K_m}/K_m)^2}$ . The estimates of  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_m$  for each peptide are approximately the same for peptide dilutions that differ by 4-fold.

Parameter	Class I	StdErr_I	Class II	StdErr_II
<b>0.44 <math>\mu\text{M}</math></b>				
$k_{\text{cat}}, \text{s}^{-1}$	6.4E-05	2.4E-06	3.6E-05	8.6E-07
$K_m, \text{M}$	0.009	0.001	0.008	0.0006
$k_{\text{cat}}/K_m, \text{s}^{-1}\text{M}^{-1}$	7.1E-03	8.3E-04	4.5E-03	3.5E-04
<b>1.75 <math>\mu\text{M}</math></b>				
$k_{\text{cat}}, \text{s}^{-1}$	1.0E-04	1.2E-06	4.30E-05	1.6E-06
$K_m, \text{M}$	0.012	0.0004	0.02	0.002
$k_{\text{cat}}/K_m, \text{s}^{-1}\text{M}^{-1}$	8.3E-03	3.0E-04	2.2E-03	2.3E-04

TABLE 5

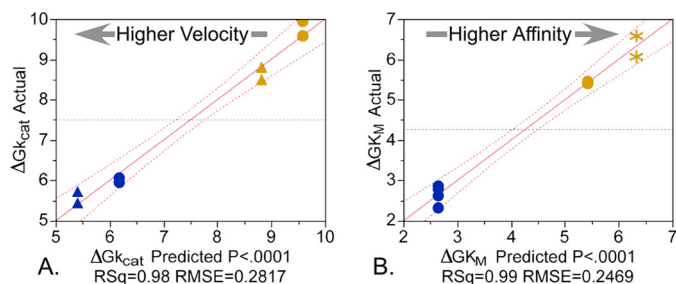
Comparison of averaged steady-state kinetic parameters of the WT and designed 46-mers

Free energies were calculated as  $\Delta G = -RT \ln(\kappa)$ , where  $\kappa$  is either  $k_{\text{cat}}$ ,  $K_m$ , or  $k_{\text{cat}}/K_m$  values are for amino acid dependence only. Standard deviations (Stdev) of the averages are given in rows below the mean free energies. A fuller comparison of individual free energies is shown in Fig. 10.

	$\langle \Delta G k_{\text{cat}} \rangle$	$\Delta G K_m$	$\Delta G k_{\text{cat}}/K_m$
	kcal/mol	kcal/mol	kcal/mol
WT	9.2	5.4	3.8
Stdev	0.7	0.6	1.0
Designed	5.8	2.7	3.2
Stdev	0.3	0.2	0.4

TyrRS) Class I and five (HisRS, ProRS, AspRS, AsnRS, and PheRS) Class II sequences as described for TrpRS, TyrRS, HisRS, and ProRS (11).

## Both Strands of a Gene Can Encode Functional Enzymes



**FIGURE 10. Significant clustering of steady-state parameters for wild-type (amber) and designed (blue) Class I and II 46-mer peptides.** Plots show observed values for the free energies of activation derived from  $k_{cat}$  and binding,  $K_m$ , versus calculated values from linear regression models using the coefficients for the binary descriptors: Designed {0,1}, Class {1(▲),2(●)} ( $k_{cat}$  only), and substrate {ATP (\*) = 1, amino acid (●) = 2} ( $K_m$  only). A, two designed 46-mers have substantially higher turnover numbers ( $\langle \Delta Gk_{cat} \rangle = 5.8$  kcal/mol) than the wild-type 46-mers ( $\langle \Delta Gk_{cat} \rangle = 9.2$  kcal/mol), and the TrpRS (Class I) wild-type 46-mer has significantly higher turnover numbers than the HisRS (Class II) 46-mer with both amino acid and ATP substrates. B, both wild-type 46-mers bind amino acid substrates with higher affinities ( $\langle \Delta GK_M \rangle = 5.9$  kcal/mol) than does either designed 46-mer ( $\langle \Delta GK_M \rangle = 2.7$  kcal/mol). Similarly, wild-type 46-mers have higher affinity for ATP than they do for amino acid substrates. Student's  $t$  test probabilities for these effects under the null hypothesis are  $< 0.0001$  for WT sequences in both models, 0.01 for Class I in the model for  $\Delta Gk_{cat}$ , and 0.01 for substrate in the model for  $\Delta GK_M$ , so these differences are statistically significant.

**TABLE 6**

**Coefficients for regression models relating  $\Delta Gk_{cat}$  and  $\Delta GK_m$  to independent design variables**

Designed is 1 for designed and 0 for WT 46-mers. Class is 1 for Class I and 2 for Class II. Substrate is 1 for ATP and 2 for amino acid.  $R^2$  values for the regression models are  $\sim 0.99$  and are given below the  $x$  axes of Fig. 10. The intercept is the constant term in the regression model and is equal to the average over all values of the dependent variable. Std. Error is standard error.

Term	Coefficient	Std. Error	$t$ ratio	Probability $>  t $
<i>kcal/mol</i>				
$\Delta Gk_{cat}$				
Intercept	8.1	0.2	34.8	$< 0.0001$
Designed	-3.4	0.1	-24.5	$< 0.0001$
Class	0.8	0.1	5.5	0.01
(Designed)*(Class)	-0.7	0.3	-2.5	0.07
$\Delta GK_m$				
Intercept	5.9	0.1	47.6	$< 0.0001$
Designed	-2.8	0.2	-13.0	$< 0.0001$
Substrate	-0.5	0.1	-3.7	0.01

Together with functional activities of the wild-type sequences described above, the differences between all three sets of sequences suggest that a relatively large number of sequences may be consistent with comparable catalytic activities for both types of peptides. It may be noteworthy that Rosetta selected consensus amino acids from the biological MSAs shown in the logos in Fig. 6B at 32% of the positions for the Class I, but at only 7% for the Class II peptide, suggesting different tolerances to mutation of the 2-folds.

### Discussion

The validity of our conclusions depends on establishing the authenticity of catalysis by wild-type and the designed 46-mer peptides described here. It is therefore important to outline arguments that the results may be artifacts. These include the following: (i) the data are very surprising; (ii) the rate accelerations are based on experimental data scarcely above the level of detection and may result from contamination by wild-type enzymes; (iii) turnover numbers for the designed 46-mers,

which were purified as fusion proteins, are much higher than those of the wild-type 46-mers, which were purified from inclusion bodies, consistent with higher concentrations of contaminating full-length enzymes in the former; (iv) high concentrations of wild-type 46-mers place upper bounds on  $K_m$  values.

The following points summarize reasons we feel that the data represent authentic catalysis by both sets of 46-mers. (i) Comparable levels of full-length enzymes should exist in both 46-mer and MBP preparations; yet MBP controls show no activity. (ii) Activities of the designed 46-mers depend on the presence of wild-type residues in positions known from studies of the full-length enzymes to be essential for full catalytic activity. (iii) Increased activity results from higher concentrations only of designed 46-mers; neither MBP nor mutant controls exhibited this behavior. (iv) Quenching of ATP fluorescence by WT 46-mers cannot be explained by minute amounts of contaminating full-length enzymes. (v) There is qualitative agreement between the relative ATP affinities measured by fluorescence quenching and those determined by steady-state kinetics for the wild-type 46-mers. (vi) Upper bounds for  $K_m$  values for ATP substrates of the wild-type 46-mers are severalfold smaller than the corresponding values for full-length enzymes. (vii)  $K_m$  values for amino acid substrates of the designed peptides are substantially increased, relative to those of the full-length enzymes. (viii) The comparison between the WT and designed 46-mers is consistent with expectations that enforcing genetic complementarity by protein design might reduce the specificities of the WT peptides by increasing  $K_m$  values for amino acid substrates.

The data reported here furnish strong, complementary sources of evidence for the following conclusions. However, the results are important enough to merit validation by further experiments. Such experiments include characterizing multiple designed sense/antisense genes from Rosetta runs sampling more of sequence space, as well as combinatorial mutations, coded by complementary codons, of active-site residues that were fixed in this design.

*Direct Experimental Validation of the Sense/Antisense Coding Hypothesis*—Catalysis by both designed 46-mers directly substantiates the sense/antisense coding ancestry of Class I and II aaRS. Together with biochemical (1, 5, 7, 8, 10) and bioinformatic (11) evidence and the repeated failure to falsify the Rodin-Ohno hypothesis (1), they are an exacting, direct experimental test of the duality proposed (4) to underlie the ancestral aaRS classes and represent an important unification in biology.

*Unique Information in a Gene Can Have Two Alternative Functional Interpretations*—Comparable activities for products of both strands of the designed 46-mer gene directly confirm that two distinct interpretations exist for the information contained in some genes. Here, remarkably, the two gene products catalyze essentially the same reaction and with approximately the same rate accelerations. Yet, the two distinct catalysts are tightly linked through strict genetic complementarity. Complementary gene sequences impose such a strong constraint that one would expect such coding to be rapidly superseded, leaving computational modeling such as we have reported here as perhaps the only access to such extinct information.

*Catalysts Like These 46-mers Were Essential for the Origin of Translation*—We argued previously (5, 8) from model reactions (14) that the uncatalyzed rate of amino acid activation is  $\sim 8.3E-9 \text{ s}^{-1} \text{ M}^{-1}$ . However, the essential time independence of MBP and mutant peptide controls imply that we have approximated the uncatalyzed rate correctly with a slightly slower rate of  $\sim 4.2E-9 \text{ s}^{-1} \text{ M}^{-1}$ . Apparent second-order rate constants for the Class I and II peptides  $\sim 6E-3 \text{ s}^{-1} \text{ M}^{-1}$  (Table 4) suggest that they accelerate amino acid activation 750,000–1,300,000-fold.

Uncatalyzed peptide bond formation from activated amino acids,  $3 \times 10^{-5} \text{ s}^{-1} \text{ M}^{-1}$  (12), is 2 orders of magnitude slower than the 46-mer catalyzed rates. Thus, the existence of similar catalytic peptides would eliminate amino acid activation as the kinetic barrier to peptide bond synthesis, even without a ribosome. Such catalysis could therefore have had a decisive and pleiotropic selective advantage, perhaps helping to select early evolutionary development of ribosomes, whose emergence might therefore have co-evolved later, with more sophisticated peptide-based aminoacyl-tRNA synthetases (e.g. synthetase Urzymes).

*Sense/Antisense 46-mers May Represent a Transition between Uncoded and Genetically Coded Peptides*—The small size of the 46-mers and their widespread conservation throughout the proteome (44) underscore the significance of their catalytic activities. They contain the Class I (PXXXXHXGH; a P-loop variant) and Class II (VTDVXXXXXR) phylogenetic signatures of longer, more sophisticated aaRS Urzymes; yet they are short enough to bridge the gap between Urzymes and simpler uncoded peptides. They might therefore be described as “Protozymes.”

*Do Amino Acid Specificities of the Two Peptides Reflect Contemporary Synthetase Class Preferences?*—Neither peptide offers a readily identifiable amino acid-binding site. However, different patterns of side chains and hydrogen bonding in parallel (Class I) versus antiparallel (Class II)  $\beta$ -structures provide a potential basis for differently sized pockets that might discriminate between sizes of otherwise structurally related amino acids, as proposed by Belrhali *et al.* (20, 45).

*Catalysis Probably Arises from Transient Structures*—The PDB files of the designed peptides output by Rosetta (supplemental material) closely resemble the native conformations that constrained them. The work of Mildvan and co-workers (28–31) suggests, however, that the predominant structures of homologous peptides in solution, and which can be determined by NMR methods, do not closely resemble the structures that the same peptides adopt in the native protein crystal structures, except in the presence of ligands. Thus, the peptide catalysts described here probably adopt transient structures in the presence of substrates that enable them to stabilize the transition states for amino acid activation. Such transient structures may actually differ markedly from what is observed in crystal structures (25, 26). That transient structures may be able to form tighter bonds to transition states has begun to be discussed in detail from both experimental (57) and theoretical (46) standpoints. In turn, this suggests that the peptides have significant normal modes allowing them to sample catalytically active conformations. Access to such structures by computational biology is thus an important and intriguing goal.

*Experimental Recapitulation of Possible Assembly of Other Modular Components of Both Synthetase Families*—Measurable 46-mer catalysis makes it possible to ask how the two additional modules of the Urzymes, related roughly to specific amino acid and tRNA binding, became assimilated by testing their effects in *cis* and in *trans*. Furthermore, complementation can be characterized in greater detail by measuring the energetic coupling between them using modular thermodynamic cycles (7, 43). Characterization of additional pairs of WT 46-mers and designed sense/antisense peptides also may help resolve the question of whether the constraint of genetic complementarity is inconsistent with high amino acid affinity.

*Class I and II 46-mers May Have Spawned an Extensive Portion of the Contemporary Proteome*—Unification of two entirely different aaRS superfamilies through their complementary genetic ancestry may signal more substantial relationships yet to be verified.

The corner between what in full-length crystal structures is the  $\alpha$ -helix and the second  $\beta$ -strand of the Class I 46-mer occurs with minimal variation of amino acid, spacing, and three-dimensional packing in  $\sim 125$  different protein families from the Rossmannoid superfamily (44), the largest in the proteome (47–49). P-loop peptides studied by Mildvan (28–31) also arguably have structural homology to parts of the Class I aaRS 46-mers that bind ATP; the segment N-terminal to the first helix shares glycines in homologous positions before, within, and following the Class I HIGH signature. The root mean square deviation = 3.6 Å for 25 residues, including the P-loop, preceding  $\beta$ -strand, and subsequent  $\alpha$ -helix, and superimposes two conserved Gly residues; see also Fig. 11A in Ref. 1.

These observations suggest that the Class I 46-mer has widespread distribution elsewhere in the proteome (44) and may thus have been ancestral not only to the Rossmannoid superfamily but also to other protein superfamilies. These include the enzymes characterized by Mildvan and co-workers (25–29) and a wider class of energy-transducing enzymes, including myosins (58), AAA<sup>+</sup> motors (50, 51), STAND regulators of programmed cell death (52), and signaling GTPases (59, 60), all of which use P-loops to activate ATP and hence are arguably directly homologous to the Class I 46-mer. Further afield, the superfamily of metabolic enzymes in nucleotide biosynthetic pathways (53, 54) may also have its origins in peptides related to the Class I 46-mer.

One also expects that the Class II 46-mer would have driven a comparable adaptive radiation, generating homologies with a comparable range of superfamilies. The comparative anatomy of antiparallel  $\beta$  proteins, however, suggests a more plastic structure than the readily recognizable parallel  $\beta$ - $\alpha$ - $\beta$  topology. We have noted previously that the motif 2 loop in the Class II 46-mer may be homologous to the nucleotide-binding sites in the actin/HSP70 superfamily (10). We note here as well that nucleotide-binding sites in the majority of nucleic acid polymerases, containing conserved aspartate  $\text{Mg}^{2+}$  ligands, have a similar  $\beta$ - $\alpha$ - $\beta$  topology to that of the Class II motif 2 signature, albeit with a generally much longer  $\alpha$ -segment.

Catalytic properties of the 46-mers also support previous suggestions (41) that the earliest catalytic peptides mobilized ATP and other NTPs for biosynthetic purposes. Thus, the pep-



## Both Strands of a Gene Can Encode Functional Enzymes

tides described here catalyze a synthetic reaction central to biology, and phylogenetic analysis suggests they may be among the most broadly conserved and hence perhaps the oldest surviving functional peptide motifs in the proteome.

**Author contributions**—C. W. C. Jr., L. L., and O. E. conceived the study. O. E., X. A., and B. K. designed the gene. L. L. isolated and V. W. performed fluorescence titrations and assayed the wild-type peptides. L. L. performed the mutagenesis of the active sites. M. C. purified the MBP used as a control. L. M.-R., K. G.-R., M. J.-R., and T. W. coordinated the purification and assay of the designed peptides. S. N. C. performed the bioinformatic analysis shown in Fig. 6. C. W. C. Jr., V. W., L. L., and T. W. wrote the paper. All authors reviewed and interpreted the results and approved the final version of the paper.

**Acknowledgments**—We thank J. Hermans for advice on the fluorescence experiments and R. Wolfenden, T. Traut, and C. Lewis for assistance with the manuscript. Completion of this work would not have been possible without the encouragement and support of L. Parise, the American Biophysical Society Summer Course, and the University of North Carolina at Chapel Hill Post-baccalaureate Research Education and Summer Undergraduate Research Experience programs.

### References

1. Carter, C. W., Jr., Li, L., Weinreb, V., Collier, M., Gonzalez-Rivera, K., Jimenez-Rodriguez, M., Erdogan, O., Kuhlman, B., Ambroggio, X., Williams, T., 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
2. Carter, C. W. (2015) What RNA world? Why a peptide/RNA partnership merits renewed experimental attention. *Life* **5**, 294–320
3. Rodin, A. S., Rodin, S. N., and Carter, C. W., Jr. (2009) On primordial sense-antisense coding. *J. Mol. Evol.* **69**, 555–567
4. Rodin, S. N., and Ohno, S. (1995) Two types of aminoacyl-tRNA synthetases could be originally encoded by complementary strands of the same nucleic acid. *Orig. Life Evol. Biosph.* **25**, 565–589
5. Pham, Y., Li, L., Kim, A., Erdogan, O., Weinreb, V., Butterfoss, G. L., 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
6. Carter, C. W., and Duax, W. L. (2002) Did tRNA synthetase classes arise on opposite strands of the same gene? *Mol. Cell* **10**, 705–708
7. 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
8. 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
9. Li, L., Francklyn, C., and Carter, C. W., Jr. (2013) Aminoacylating urzymes challenge the RNA world hypothesis. *J. Biol. Chem.* **288**, 26856–26863
10. Carter, C. W., Jr. (2014) Urzymology: experimental access to a key transition in the appearance of enzymes. *J. Biol. Chem.* **289**, 30213–30220
11. Chandrasekharan, S. N., Yardimci, G. G., Erdogan, O., Roach, J., and Carter, C. W., Jr. (2013) Statistical evaluation of the Rodin-Ohno hypothesis: sense/antisense coding of ancestral class I and II aminoacyl-tRNA synthetases. *Mol. Biol. Evol.* **30**, 1588–1604
12. Schroeder, G. K., and Wolfenden, R. (2007) The rate enhancement produced by the ribosome: an improved model. *Biochemistry* **46**, 4037–4044
13. Sievers, A., Beringer, M., Rodnina, M. V., and Wolfenden, R. (2004) The ribosome as an entropy trap. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7897–7901
14. Kirby, A. J., and Younas, M. (1970) The reactivity of phosphate esters. Reactions of diesters with nucleophiles. *J. Chem. Soc.* **B** **418**, 1165–1172
15. Wolfenden, R. (2011) Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu. Rev. Biochem.* **80**, 645–667
16. Stockbridge, R. B., and Wolfenden, R. (2009) The intrinsic reactivity of ATP and the catalytic proficiencies of kinases acting on glucose, *N*-acetylgalactosamine, and homoserine: a thermodynamic analysis. *J. Biol. Chem.* **284**, 22747–22757
17. Retailleau, P., Huang, X., Yin, Y., Hu, M., Weinreb, V., Vachette, P., Vonrhein, C., Bricogne, G., Roversi, P., Ilyin, V., and Carter, C. W., Jr. (2003) Interconversion of ATP binding and conformational free energies by tryptophanyl-tRNA synthetase: structures of ATP bound to open and closed, pre-transition conformations. *J. Mol. Biol.* **325**, 39–63
18. Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J.-C., and Moras, D. (1994) The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. *EMBO J.* **13**, 327–337
19. 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
20. Belrhali, H., Yaremchuk, A., Tuko, M., Berthet-Colominas, C., Rasmussen, B., Bösecke, P., Diat, O., and Cusack, S. (1995) The structural basis for seryl-adenylate and Ap4A synthesis by seryl-tRNA synthetase. *Structure* **3**, 341–352
21. Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., and Leberman, R. (1990) A second class of synthetase structure revealed by x-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347**, 249–255
22. Chan, K. W., and Koeppe, R. E., 2nd (1994) Role of the TIGN sequence in *E. coli* tryptophanyl-tRNA synthetase. *Biochim. Biophys. Acta* **1205**, 223–229
23. Fersht, A. R. (1987) Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. *Biochemistry* **26**, 8031–8037
24. Arnez, J. G., Augustine, J. G., Moras, D., and Francklyn, C. S. (1997) The first step of aminoacylation at the atomic level in histidyl-tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7144–7149
25. Chuang, W.-J., Abeygunawardana, C., Pedersen, P. L., and Mildvan, A. S. (1992) Two-dimensional NMR, circular dichroism, and fluorescence studies of PP-50, a synthetic ATP-binding peptide from the b-subunit of mitochondrial ATP synthase. *Biochemistry* **31**, 7915–7921
26. Chuang, W.-J., Abeygunawardana, C., Gittis, A. G., Pedersen, P. L., and Mildvan, A. S. (1995) Solution structure and function in trifluoroethanol of PP-50, an ATP-binding peptide from F<sub>1</sub>ATPase. *Arch. Biochem. Biophys.* **319**, 110–122
27. Mullen, G. P., Vaughn, J. B., Jr., and Mildvan, A. S. (1993) Sequential proton NMR resonance assignments, circular dichroism, and structural properties of a 50-residue substrate-binding peptide from DNA polymerase I. *Arch. Biochem. Biophys.* **301**, 174–183
28. Fry, D. C., Byler, D. M., Susi, H., Brown, E. M., Kuby, S. A., and Mildvan, A. S. (1988) Solution structure of the 45-residue MgATP-binding peptide of adenylate kinase as examined by 2-D NMR, FTIR, and CD spectroscopy. *Biochemistry* **27**, 3588–3598
29. Fry, D. C., Kuby, S. A., and Mildvan, A. S. (1985) NMR studies of the MgATP-binding site of adenylate kinase and of a 45-residue peptide fragment of the enzyme. *Biochemistry* **24**, 4680–4694
30. Ambroggio, X. I., and Kuhlman, B. (2006) Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* **128**, 1154–1161
31. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368
32. Liu, Y., and Kuhlman, B. (2006) RosettaDesign server for protein design. *Nucleic Acids Res.* **34**, 235–238
33. Leaver-Fay, A., Jacak, R., Stranges, P. B., and Kuhlman, B. (2011) A generic program for multistate protein design. *PLoS ONE* **6**, e20937
34. Kuhlman, B., and Baker, D. (2004) Exploring folding free energy land-

- scapes using computational protein design. *Curr. Opin. Struct. Biol.* **14**, 89–95
35. SAS Institute (2007) *JMP Statistics and Graphics Guide*, Version 6 Ed., SAS Institute, Cary, NC
  36. Hardin, R. H., and Sloane, N. J. (1993) A new approach to the construction of optimal designs. *J. Stat. Plan. Inf.* **37**, 339–369
  37. Fersht, A. R. (1999) *Structure and Mechanism in Protein Science*, pp. 89–95, W. H. Freeman and Co., New York
  38. Box, G. E., Hunter, W. G., and Hunter, J. S. (1978) *Statistics for Experimenters*, Wiley Interscience, New York
  39. Weinreb, V., Li, L., and Carter, C. W., Jr. (2012) A master switch couples  $Mg^{2+}$ -assisted catalysis to domain motion in *B. stearothermophilus* tryptophanyl-tRNA synthetase. *Structure* **20**, 128–138
  40. Augustine, J., and Francklyn, C. (1997) Design of an active fragment of a class II aminoacyl-tRNA synthetase and its significance for synthetase evolution. *Biochemistry* **36**, 3473–3482
  41. Härtle, M., and Cusack, S. (1995) Structure, function and evolution of seryl-tRNA synthetases: implications for the evolution of aminoacyl-tRNA synthetases and the genetic code. *J. Mol. Evol.* **40**, 519–530
  42. 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
  43. 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
  44. 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
  45. Belrhali, H. (1996) *Determination by X-ray Crystallography of the Mechanisms by Which Seryl-Adenylate and bis(5'-Adenosyl)-Tetraphosphate Are Formed by the Seryl-tRNA Synthetase from Thermus thermophilus*. Thèse d'état, European Synchrotron Radiation Facility
  46. Hu, H. (2014) Wild-type and molten globular chorismate mutase achieve comparable catalytic rates using very different enthalpy/entropy compensations. *Sci. China Chem.* **57**, 156–164
  47. Dokholyan, N. V., Shakhnovich, B., and Shakhnovich, E. I. (2002) Expanding protein universe and its origin from the biological big bang. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14132–14136
  48. Dokholyan, N. V., and Shakhnovich, E. I. (2001) Understanding hierarchical protein evolution from first principles. *J. Mol. Biol.* **312**, 289–307
  49. Koonin, E. V. (2007) The biological big bang model for the major transitions in evolution. *Biol. Direct* **2**, 21
  50. Snider, J., Thibault, G., and Houry, W. A. (2008) The AAA<sup>+</sup> superfamily of functionally diverse proteins. *Genome Biol.* **9**, 216
  51. Iyer, L. M., Leipe, D. D., Koonin, E. V., and Aravind, L. (2004) Evolutionary history and higher order classification of AAA<sup>+</sup> ATPases. *J. Struct. Biol.* **146**, 11–31
  52. Leipe, D. D., Koonin, E. V., and Aravind, L. (2004) STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28
  53. Armenta-Medina, D., Segovia, L., and Perez-Rueda, E. (2014) Comparative genomics of nucleotide metabolism: a tour to the past of the three cellular domains of life. *BMC Genomics* **15**, 800
  54. Caetano-Anolles, G., Kim, H. S., and Mittenthal, J. E. (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9358–9363
  55. Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190
  56. Cleland, W. W., and Northrop, D. B. (1999) Energetics of substrate binding, catalysis, and product release. *Methods Enzymol.* **308**, 3–48
  57. Pervushin, K., Vamvaca, K., Vögeli, B., and Hilvert, D. (2007) Structure and dynamics of a molten globular enzyme. *Nat. Struct. Mol. Biol.* **14**, 1202–1206
  58. Sweeney, H. L., and Houdusse, A. (2010) Structural and functional insights into the Myosin motor mechanism. *Annu. Rev. Biophys.* **39**, 539–557
  59. Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **317**, 41–72
  60. Mildvan, A. S. (1997) Mechanisms of signaling and related enzymes. *Proteins* **29**, 401–416