

# Aminoacylating Urzymes Challenge the RNA World Hypothesis\*<sup>♦</sup>

Received for publication, June 22, 2013, and in revised form, July 17, 2013. Published, JBC Papers in Press, July 18, 2013, DOI 10.1074/jbc.M113.496125

Li Li<sup>‡</sup>, Christopher Francklyn<sup>§</sup>, and Charles W. Carter, Jr.<sup>‡1</sup>

From the <sup>‡</sup>Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599-7260 and the <sup>§</sup>Department of Biochemistry, University of Vermont, Burlington, Vermont 05405

**Background:** RNA World scenarios require high initial fidelity, greatly slowing lift-off.

**Results:** Class I TrpRS and Class II HisRS Urzymes (120–130 residues) both acylate tRNAs  $\sim 10^6$  times faster than the uncatalyzed peptide synthesis rate.

**Conclusion:** Urzymes appear highly evolved, implying that they had even simpler ancestors.

**Significance:** High Urzyme catalytic proficiencies imply that translation began in a Peptide-RNA World.

We describe experimental evidence that ancestral peptide catalysts substantially accelerated development of genetic coding. Structurally invariant 120–130-residue Urzymes (Ur = primitive plus enzyme) derived from Class I and Class II aminoacyl-tRNA synthetases (aaRSs) acylate tRNA far faster than the uncatalyzed rate of nonribosomal peptide bond formation from activated amino acids. These new data allow us to demonstrate statistically indistinguishable catalytic profiles for Class I and II aaRSs in both amino acid activation and tRNA acylation, over a time period extending to well before the assembly of full-length enzymes and even further before the Last Universal Common Ancestor. Both Urzymes also exhibit  $\sim 60\%$  of the contemporary catalytic proficiencies. Moreover, they are linked by ancestral sense/antisense genetic coding, and their evident modularities suggest descent from even simpler ancestral pairs also coded by opposite strands of the same gene. Thus, aaRS Urzymes substantially pre-date modern aaRS but are, nevertheless, highly evolved. Their unexpectedly advanced catalytic repertoires, sense/antisense coding, and ancestral modularities imply considerable prior protein-tRNA co-evolution. Further, unlike ribozymes that motivated the RNA World hypothesis, Class I and II Urzyme-tRNA pairs represent consensus ancestral forms sufficient for codon-directed synthesis of nonrandom peptides. By tracing aaRS catalytic activities back to simpler ancestral peptides, we demonstrate key steps for a simpler and hence more probable peptide-RNA development of rapid coding systems matching amino acids with anticodon trinucleotides.

The RNA World Hypothesis (1, 2) holds that RNA molecules, simultaneously playing both informational and catalytic roles, evolved extensively prior to the appearance of proteins. However, no biological ribozymes that might support the hypothesis participate in either replication or tRNA aminoacylation. Moreover, Koonin estimates that assembling the repli-

cative fidelity necessary for an RNA-only origin would require multiple universes (3). tRNA aminoacylation, the defining reaction in codon-dependent translation, has not previously been demonstrated for aaRS<sup>2</sup> fragments smaller than intact catalytic domains (4, 5). Experimental acylation of mini- and microhelicases derived from tRNAs (6, 7) has been shown, however, for several intact aaRSs, suggesting that aaRS catalytic domains might have acylated cognate tRNAs using an earlier “operational RNA code” (8) focused on bases in the 3' acceptor stems. We previously showed that ancestral peptides afford a realistic alternative to ribozymal catalysis by characterizing 120–130-residue Urzymes from both classes (Fig. 1) that accelerate cognate amino acid activation by ATP  $\sim 10^8$ -fold (9–11).

However, to have helped implement genetic coding (8), Urzyme-like peptide catalysts must also have acylated cognate tRNAs. Our previous work did not address that question. We demonstrate here that Class I TrpRS and Class II HisRS Urzymes indeed catalyze tRNA acylation at a far faster rate than is required for spontaneous, ribosome-independent peptide synthesis (12). These new data challenge the notion that genetic coding was established entirely by RNA molecules.

## EXPERIMENTAL PROCEDURES

TrpRS (9) and HisRS (11) Urzymes were expressed and purified as described. *Bacillus stearothermophilus* tRNA<sup>Trp</sup> and *Escherichia coli* tRNA<sup>His</sup> *in vitro* transcripts were labeled with <sup>32</sup>P at the 3'-terminal internucleotide linkage using the exchange reaction catalyzed by *E. coli* tRNA nucleotidyltransferase, as described (13). Labeled tRNA was purified on a Bio-Rad P30 spin column and then mixed with unlabeled tRNA to the desired final concentration. tRNA was refolded by heating to 80 °C for 2–3 min followed by the addition of MgCl<sub>2</sub> to a final concentration of 10 mM and slow cooling to ambient temperature over a period of 15–30 min. Aminoacylation reaction mixtures contained 50 mM HEPES, pH 7.5, 20 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM DTT, 10 mM ATP, 250 μM amino acid, enzyme (5 nM wild type; 1–5 μM Urzyme), and a range of tRNA concen-

\* This work was supported by NIGMS/National Institutes of Health Grants 78227 (to C. W. C.) and 54899 (to C. F.).

♦ This article was selected as a Paper of the Week.

<sup>1</sup> To whom correspondence should be addressed: Dept. of Biochemistry and Biophysics, CB 7260, University of North Carolina, Chapel Hill, NC 27599-7260. Tel.: 919-966-3263; Fax: 919-966-2852; E-mail: carter@med.unc.edu.

<sup>2</sup> The abbreviations used are: aaRS, aminoacyl-tRNA synthetase;  $\Delta\Delta G^\ddagger$ , TS stabilization; LUCA, Last Universal Common Ancestor; PDB, Protein Data Bank; TS, transition state.

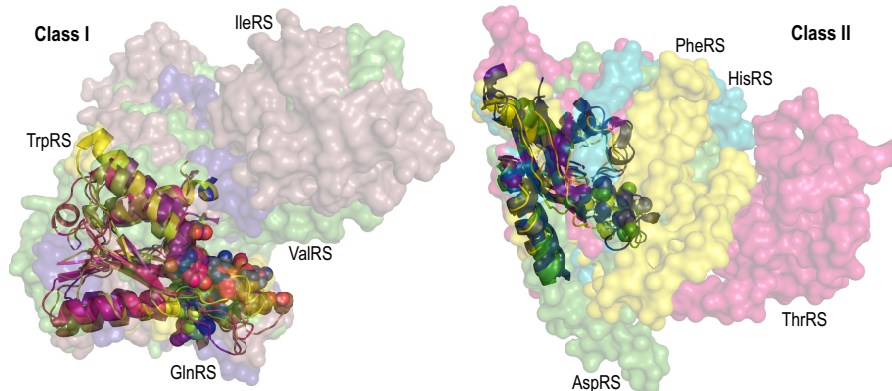


FIGURE 1. Superposition of four Class I and four Class II aaRS catalytic domains (65% transparent). Invariant cores (120–130 residues) are shown as ribbons inside differently colored surfaces, and active site residues are shown in CPK spheres.

TABLE 1

Steady-state kinetic constants from cognate amino acid-dependent  $P_i$  exchange and cognate tRNA acylation by TrpRS and HisRS Urzyme constructs

$k_{\text{cat}}$  values were normalized based on concentration of Urzyme enzyme active sites.

Catalyst	$k_{\text{cat}}$ $s^{-1}$	$K_m$ (amino acid/tRNA) $\mu M$	$k_{\text{uncat}}; k_{\text{cat}}/K_m$ $s^{-1}M^{-1}$	$\Delta G^\ddagger(k_{\text{cat}}/K_m)$ $kcal/mol$
<b>Activation<sup>a,b</sup> (Uncat)</b>				
<b>TrpRS</b>				
Urz (W)	$8.9 \pm 3.1E-04$	$2400 \pm 1200$	$3.7E-01$	0.6
Urz (ATP)	$3.7 \pm 2.3E-04$	$240 \pm 230$	$1.5E+00$	-0.3
WT_Trp	$1.60 \pm 0.11$	$3.2 \pm 1.4$	$5.5E+05$	-7.9
WT_ATP	$5.14 \pm 0.44$	$300 \pm 230$	$1.8E+04$	-5.9
<b>HisRS</b>				
Urz-3 (H)	$4.7 \pm 0.4E-04$	$260 \pm 90$	$1.8E+00$	-0.3
Urz-3 (ATP)	$2.4 \pm 0.4E-04$	$11 \pm 6$	$2.2E+01$	-1.8
WT_His	$142 \pm 5$	$30 \pm 5$	$4.7E+06$	-9.2
WT_ATP	$120 \pm 4$	$890 \pm 600$	$1.4E+05$	-7.1
<b>Acylation<sup>c</sup> (Uncat)</b>				
<b>TrpRS</b>				
Urz (tRNA)	$1.7 \pm 1.3E-03$	$7.49 \pm 0.9$	$3.0E+02$	-3.4
WT	$0.50 \pm 0.11$	$0.57 \pm 0.17$	$8.9E+05$	-8.2
<i>E. coli</i> <sup>d</sup>	26.00	0.30	$2.0E+06$	-8.7
<b>HisRS</b>				
<HisRS1-4>	$5.1 \pm 1.3E-04$	$4.9 \pm 1.6$	$1.1E+02$	-2.8
WT	$1.25 \pm 0.15$	$4.9 \pm 1.2$	$2.6E+05$	-7.4
WT <sup>e</sup>	$2.6 \pm 0.4$	$1.4 \pm 0.4$	$1.9E+06$	-8.6
<b>Peptidyl-transfer<sup>f</sup> (Uncat)</b>				
			$3.00E-05$	6.2

<sup>a</sup> Assayed by  $^{32}P_i$  exchange.

<sup>b</sup> Ref. 16.

<sup>c</sup> Ref. 17.

<sup>d</sup> Ref. 42.

<sup>e</sup> Ref. 4.

<sup>f</sup> Ref. 12.

trations (0, 0.5, 1.0, 2.5, 5.0, 10.0, and 15.0  $\mu M$ ). Reactions were run at 37 °C, for variable lengths of time, determined to be in the linear range, and 1- $\mu l$  aliquots were removed at varying time points, quenched in 400 mM sodium acetate, pH 5.2. Aminoacylated tRNA was digested with 0.1 mM P1 nuclease (Sigma), spotted on prewashed polyethyleneimine-cellulose TLC plates (Sigma), and developed in 100 mM ammonium acetate, 5% acetic acid. Dried TLC plates were quantified by phosphor imaging analysis. TLC profiles afford quantitative estimates for the amounts of acylated and unacylated A76 bases. The acylated product produced at a given time is thus the product of the initial tRNA concentration times the fraction of acylated A76. Steady-state kinetic parameters were determined using JMP (14) and corrected for the fraction of active enzyme (0.35–0.7) and aminoacylatable tRNA (0.2–0.55).

Experimental differences between Urzyme and full-length catalytic rate accelerations, between amino acid activation and acylation rates, and between Class I- and Class II-derived Urzymes were compared quantitatively by fitting linear regression models for transition-state stabilization free energies. Activation free energies for the uncatalyzed rates of amino acid activation and acyl-transfer from Table 1 were subtracted from corresponding experimental values for the catalyzed rates,  $\Delta G^\ddagger k_{\text{cat}}/K_m$ , giving estimates for transition state (TS) stabilization,  $\Delta\Delta G^\ddagger$ . TS stabilization energies are then assigned binary codes (1 or 0), according to which enzyme and which reaction they belong to, whether the catalyst includes the Urzyme and whether it includes enhancements found in the full-length enzymes. TS stabilization energies are collected with these codes in Table 2.

# Aminoacylating Urzymes Challenge the RNA World

**TABLE 2**

Transition state stabilization free energies for reactions catalyzed by TrpRS and HisRS catalysts

Trp (1,0) implies catalyst is derived from TrpRS, HisRS; Act, reaction catalyzed is amino acid activation; Urz, catalyst contains the Urzyme; Native enhancement, catalyst is the full-length enzyme.

Catalyst	TS stabilization	Trp	Act	Urz	Native enhancement
	<i>kcal/mol</i>				
Trp Urzyme	-10.25	1	1	1	0
Trp Urzyme	-8.9	1	0	1	0
His Urzyme	-11.45	0	1	1	1
His Urzyme	-8.3	0	0	1	1
Trp Native	-17.3	1	1	1	0
Trp Native	-13.95	1	0	1	1
His Native	-18.55	0	1	1	1
His Native	-13.5	0	1	1	
Uncat	0	0	1	0	0
Uncat	0	0	0	0	0

**TABLE 3**

Regression model describing relative contributions to transition state stabilization by Class I and II catalysts for amino acid activation and tRNA aminoacylation

$R^2$  for this model is 0.99, and the root mean square error is 0.82 kcal/mole for 10 observations, 5 of which are degrees of freedom for determining the error.

Term	Estimate	S.E.	<i>t</i> ratio	Prob>  <i>t</i>
Intercept	1.3	0.63	1.98	0.1047
Urz	-9.7	0.71	-19.09	<0.0001
Native enhancement	-6.1	0.58	-10.58	0.0001
Act	-2.6	0.52	-4.85	0.0047
(Native enhancement)*(Act)	-2.7	1.05	-2.53	0.0523

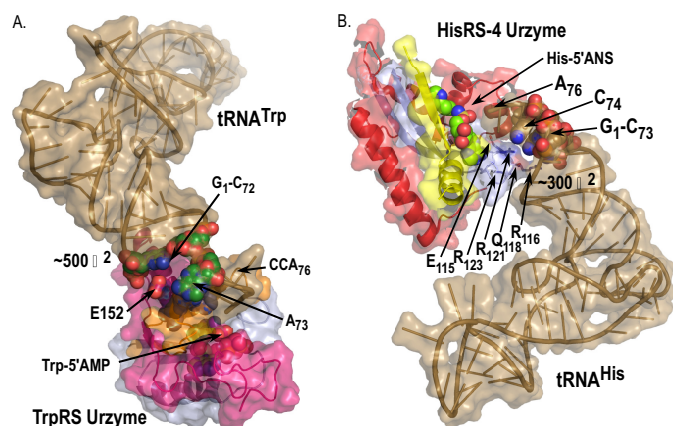
The codes in Table 2 are potential predictors of catalytic rate enhancement. The presence of the Urzyme in a catalyst may contribute consistently to the observed rate enhancement; the activation reaction may be accelerated consistently by a greater amount than tRNA aminoacylation; enhancements found in contemporary enzymes but not in Urzymes may accelerate rates by a consistent amount, and Class I catalysts may be consistently faster or slower than Class II catalysts. The extent to which statement is actually validated by the data can be assessed from the coefficients,  $\beta_i$ ,  $\beta_i\beta_j$ , of the linear model for transition state stabilization energy,

$$\Delta\Delta G^\ddagger = \text{Constant} + \sum\beta_i X_i + \sum\beta_{ij} X_i X_j \quad (\text{Eq. 1})$$

where  $X_i$  and  $X_j$  represent entries in the third through sixth columns of Table 2. This treatment differs from principal components analysis in that we explicitly postulate the identity of the principal components, rather than identifying them *a priori*. It represents multidimensional thermodynamic cycles (see Fig. 5A) These calculations were also performed with JMP (14). Validity can be evaluated from the degree to which it accounts for the variance of the TS stabilization ( $R^2 = 0.99$ ) and from Student's *t* test probabilities of the coefficients, Table 3. The four predictors in Table 3 account for virtually all of the variance in the observed values (see Fig. 5B); errors are defined with five degrees of freedom, and Student's *t* test probabilities for all but the activation\*full-length enhancements interaction term are highly significant, as discussed previously (11, 15).

## RESULTS AND DISCUSSION

TrpRS and HisRS Urzymes both retain binding sites for cognate tRNA acceptor stems. Models of aaRS-tRNA complexes



**FIGURE 2. Models for tRNA TrpRS, HisRS Urzyme complexes, derived from crystal structures of intact components.** Orientations are similar to those in Fig. 1. A, TrpRS. Model shows the putative Urzyme-tRNA<sup>Trp</sup> complex, derived from the complex between human TrpRS and tRNA<sup>Trp</sup> (PDB ID 2AZX) and that of *Bacillus stearothermophilus* TrpRS (PDB ID 1MAU). The N- and C-terminal  $\beta$ - $\alpha$ - $\beta$  crossovers are colored sky blue and hot pink, respectively. The specificity helix is orange. tRNA acceptor stem determinants and matching active-site side chains are indicated. B, HisRS. Model for interaction with tRNA<sup>His</sup> was derived from the HisRS crystal structure (PDB ID 2EL9) together with a *Rhodospirillum rubrum* ProRS-tRNA<sup>Pro</sup> complex (S. Cusack and T. Crepin, unpublished observation). Motifs 1, 2, and 3 are hot pink, sky blue, and yellow, respectively. Active-site side chains (Glu-115, Arg-116, Gln-118, Arg-121, and Arg-123) that interact with RNA acceptor stem determinants are shown as sticks.

suggest that aaRS Urzymes of both classes retain interactions with cognate tRNA acceptor stems (Fig. 2). Surface area calculations show that complex formation with tRNA buries ~300–500 Å<sup>2</sup>, positioning the acceptor stems for acylation and aligning previously identified active site and acceptor stem recognition elements. Urzymes HisRS-1, HisRS-2, HisRS-3, and HisRS-4 differ in the presence (HisRS-3, HisRS-4) or absence of signature Motif 3, and the presence (HisRS-2, HisRS-4) of a 6-residue N-terminal extension preceding Motif 1 (11). All four HisRS Urzymes contain the identity elements shown in Fig. 2B. Thus, such recognition potentially confers on both Urzymes the rudimentary acceptor-stem discrimination necessary to interpret an operational RNA code (8).

*TrpRS and HisRS Urzymes Accelerate tRNA Aminoacylation ~10<sup>6</sup>-Fold*—Assays using <sup>32</sup>P-3' adenosine-labeled cognate tRNA (13) (Fig. 3) readily allowed derivation of steady-state kinetic parameters, documenting that both Urzymes catalyze cognate tRNA aminoacylation. Notably, whereas previously published second-order rate constants ( $k_{\text{cat}}/K_m$ ) for amino acid activation were decreased 10<sup>5</sup>-fold, those for aminoacylation are decreased by only three logs relative to the full-length enzymes (Table 1). Moreover, despite lacking anticodon binding domains, Urzyme tRNA  $K_m$  values are within an order of magnitude of those for contemporary WT enzymes.

Model bio-organic reactions (12, 16, 17) summarized in Table 1 provide plausible estimates of the uncatalyzed rates of amino acid activation, acyl transfer, and peptidyl-transfer (denoted by "Uncat" in the Table). These rates likely constrained the evolution of living systems prior to the advent of biological catalysts. Contemporary enzymes accelerate essential reactions differentially so that they can all proceed at approximately the same rate.

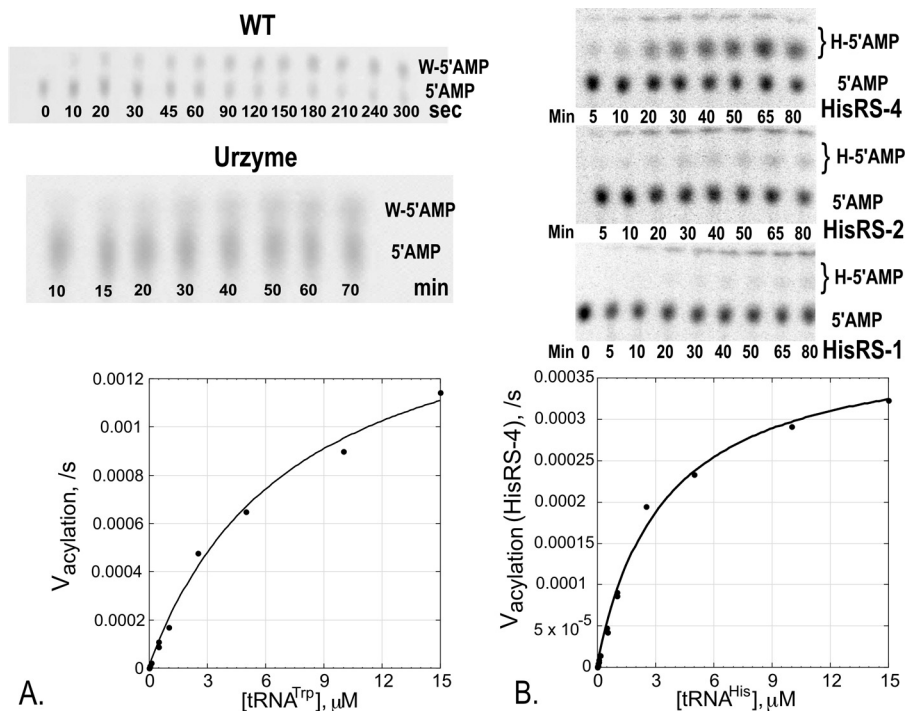


FIGURE 3. **tRNA Acylation by TrpRS (A) and HisRS (B) Urzymes.** Autoradiograms document aminoacylation of cognate tRNAs by TrpRS Urzyme (A) and HisRS-1, HisRS-2, and HisRS-4 (11) Urzymes (B). WT HisRS and HisRS-3 are similar and omitted to save space. Michaelis-Menten plots document saturation of with respective cognate tRNAs. The thin layer chromatographic assay is described in Ref. 13. Standard errors are given in Table 1.

All three reactions necessary for protein synthesis therefore must have proceeded at comparable rates for protein synthesis to evolve. The significant differences among uncatalyzed rates of these three reactions mean that efficient protein synthesis demanded greater acceleration of amino acid activation than of tRNA aminoacylation. Moreover, to preserve uniformity of translation across a variety of messages, different aaRS classes must have accumulated parallel evolutionary rate enhancements.

*Class I and II Aminoacyl-tRNA Synthetase Catalytic Profiles Evolved in Parallel*—Experimental catalyzed rates (Table 1 and Figs. 4 and 5) quantitatively confirm both expectations outlined in the previous section. The new aminoacylation data allow quantitative assessment of catalytic proficiencies of Class I and II Urzymes and contemporary enzymes in both amino acid activation and tRNA acylation. The uncatalyzed rate of amino acid activation is  $\sim 3$  orders of magnitude slower than the other two steps (Fig. 4A); hence, it is the most significant of the three chemical barriers to protein synthesis.

Class I and II Urzymes accelerate amino acid activation  $10^8$ – $10^9$ -fold (Fig. 4B). Previously, we demonstrated the authenticity of this catalytic activity (9–11), showing that the observed activity arises from major percentages of molecules, that amino acid  $K_m$  values are increased relative to those of full-length enzymes and that activities are modified by mutation. Evolutionary enhancement of both families, to create catalytic domains and then full-length enzymes, led to equivalent catalytic rate increases for both amino acid activation (Fig. 4B) and tRNA aminoacylation.

Regression modeling (Fig. 5) affords a convenient assessment of the statistical significance of patterns evident in averages of different groupings within the data. It reinterprets the TS sta-

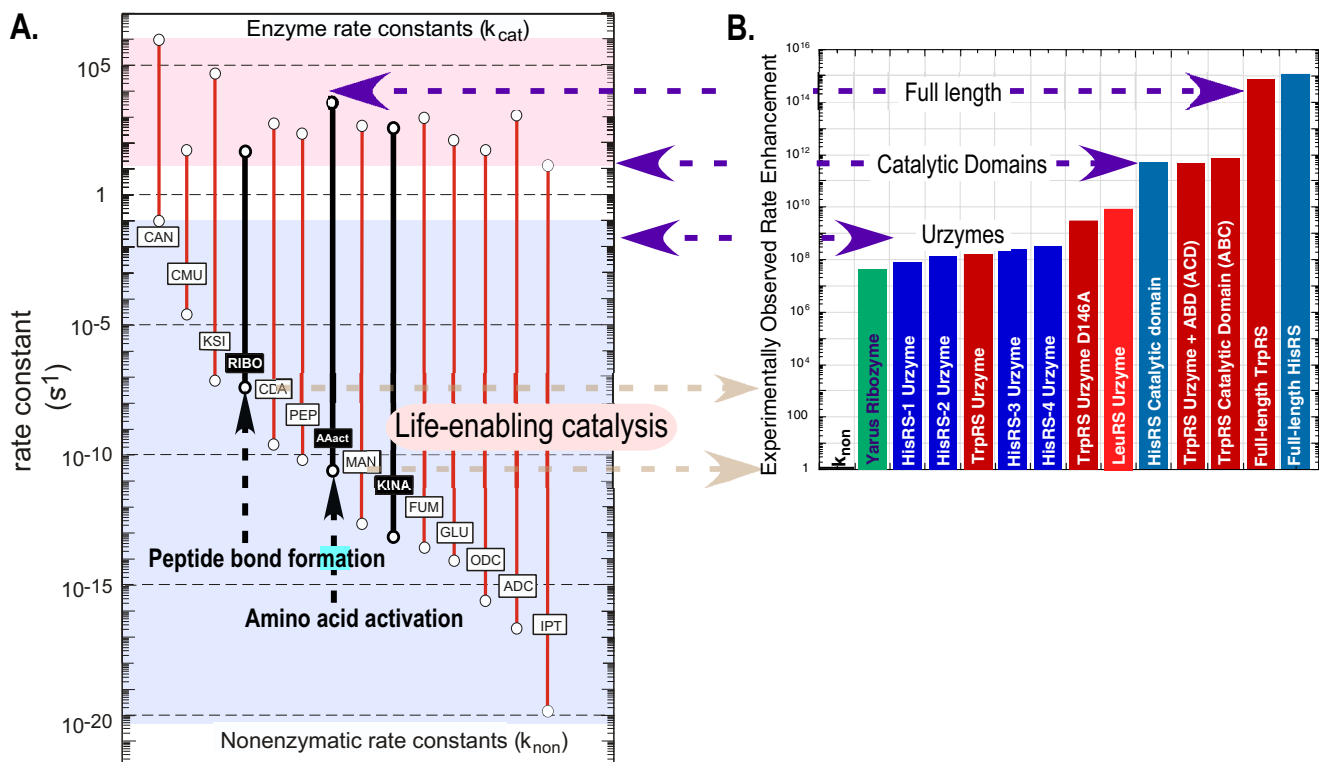
bilization values in the space spanned by the four predictor columns in Table 2 and does not introduce new information. The analysis (Table 3 and Fig. 5C) confirms that both Urzymes reduce transition state free energies of both reactions by an average  $-9.7 \pm 0.7 k_{\text{cal}}/\text{mol}$  ( $p < 0.0001$ ).

Catalytic enhancements in full-length HisRS and TrpRS enzymes further stabilize all transition states by  $-6.1 \pm 0.6 k_{\text{cal}}/\text{mol}$  ( $p < 0.0001$ ). The Urzymes therefore have only 15–25% of the mass but exhibit 61% of the catalytic proficiency of contemporary aaRSs. Quantitative comparison suggests that the Urzymes, which comprise the oldest modules of both synthetase superfamilies, are to the full-length contemporary enzymes approximately what the chassis and drive train are to an automobile.

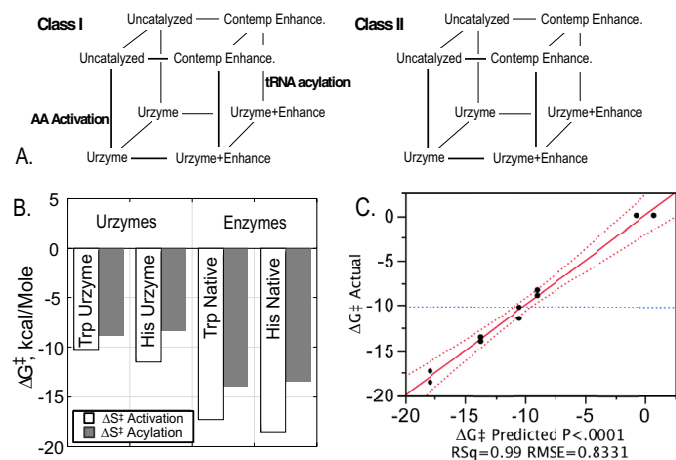
The Urzymes accelerate  $^{32}\text{P}_i$  exchange, a surrogate for cognate amino acid activation, and tRNA aminoacylation in approximately the same proportion as do the full-length enzymes. Transition states for amino acid activation are preferentially stabilized in all four cases, by  $-2.6 \pm 0.6 k_{\text{cal}}/\text{mol}$  more than those for acylation ( $p = 0.0045$ ), consistent with the fact that uncatalyzed amino acid activation is slower. Thus, all four catalysts tend to equalize the two reaction rates by preferentially accelerating amino acid activation. Both Urzymes have essentially modern catalytic profiles.

*Class I and II Urzyme-like aaRSs Date from Substantially Earlier than Modern aaRS, Last Universal Common Ancestor (LUCA)* (Fig. 6)—Consensus (18, 19) holds that protein synthesis in the LUCA was supported by a nearly complete set of essentially modern aaRSs, with fully developed accessory domains for anticodon recognition and editing of incorrectly activated amino acids. In contrast, the TrpRS and HisRS

## Aminoacylating Urzymes Challenge the RNA World



**FIGURE 4. Rates of  $^{32}\text{P}$  exchange with cognate amino acids by Class I (TrpRS, LeuRS) and Class II (HisRS) aaRS Urzymes, catalytic domains, and native enzymes, compared with uncatalyzed and fully catalyzed rates.** *A*, rate accelerations estimated from experimental data for single substrate (red) and bi-substrate (black, bold) reactions adapted to include uncatalyzed and catalyzed rates of bi-substrate reactions of the ribosome (12), amino acid activation (16), and kinases (40). Second-order rate constants (black bars) were converted into comparable units by multiplying by 0.002 M, which is the ATP concentration used to assay the catalysts shown in *B*. *B*, experimental rate accelerations estimated from steady-state kinetics as  $k_{\text{cat}}/K_m$  for a series of catalysts derived from Class I and Class II aaRSs (5, 6, 11). Vertical scales in *A* and *B* are the same. Histograms in *B* were normalized by subtracting the logarithm of the uncatalyzed rate of amino acid activation (AAact; *A*). Red bars denote Class I tryptophanyl- and leucyl-tRNA synthetase constructs, blue bars denote Class II histidyl-tRNA synthetase constructs, and green bars denote the ribosomal catalyst (18) for comparison. LeuRS Urzyme is the work of M. Collier, and the TrpRS catalytic domains made by recombining the TrpRS Urzyme with either CP1 or with the anticodon-binding domain are the work of L. Li (both unpublished observations).



**FIGURE 5. Quantitative comparison of relative transition state stabilization energies.** *A*, comparisons outlined under “Experimental Procedures” represented as multidimensional thermodynamic cycles. Edges represent contributions to the overall rate acceleration that can be attributed to each predictor in third through sixth columns of Table 2. *B*, transition state stabilization free energies from steady-state  $k_{\text{cat}}/K_m$  values for activation (open) and acylation (shaded) by Urzymes and full-length contemporary enzymes. *C*, regression model relating transition state stabilization free energies to contributions from Urzymes themselves, evolutionary enhancements in the full-length enzymes, whether the reaction is amino acid activation or tRNA acylation, and the two-way interaction between enhancements and activation (Table 2). The aaRS Class distinction does not contribute significantly.

Urzymes consist exclusively of the invariant structural cores of their respective superfamilies.

To the extent that we have characterized them, their amino acid specificities are quite modest (9). Whereas Urzymes greatly reduce the structural complexity necessary to implement translation of a rudimentary genetic code, that limitation would have prevented them from translating modern proteins. Thus, the Urzymes exhibit catalytic properties from an era well before the full expression of the genetic code and hence from a much earlier evolutionary stage than modern aaRSs or LUCA.

In light of their limited specificity, the exquisitely matched catalytic repertoires of both Urzymes may initially appear puzzling. Conservation of domains and catalytic residues involved in amino acid activation by aaRSs and variability of those required for tRNA acylation have led others (20, 21) to argue that evolution of amino acid activation activity preceded acquisition of the capability to catalyze acylation. Building on their catalytic cores, Urzymes could subsequently have accumulated anticodon binding and insertion domains in response to evolutionary selective pressure to enhance amino acid and tRNA specificity as the genetic code expanded during later stages of evolution. Integrating these new modules likely required mechanistic changes to exploit and tune conformational cycles in multidomain aaRSs (15), selecting for more recent catalytically important residues stabilizing transition states for tRNA acyla-

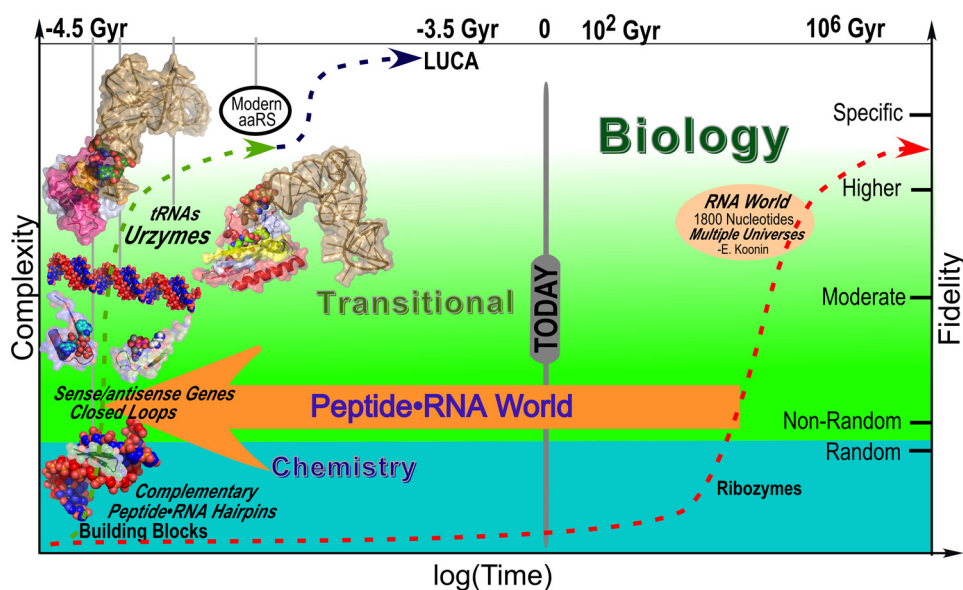


FIGURE 6. Schematic diagram of likely peptide participation with RNA in the origin and evolution of codon-directed protein synthesis (adapted from Ref. 3). Although not required, we assume that the two vertical axes are more or less aligned, that increases in fidelity are accompanied by increases in information, and *vice versa*, for schematic simplicity. The consensus (18, 19) order of appearance of modern aaRSs, tRNAs, and LUCA is reinforced by the quantitative analysis of rate data summarized in Table 2, Fig. 3, and Fig. 4. Prior existence of sense/antisense genes coding for catalytically active Class I and II closed loops (41) is based on circumstantial evidence discussed under "Results and Discussion." Initiation of a peptide-RNA World by complementary peptide and RNA hairpins follows a proposal by Carter and Kraut (36).

tion (21). These questions notwithstanding, uniform integration of both activities into the same Urzymes is consistent with their participation in implementing an operational RNA code (8).

*Class I and II Urzymes Are Highly Evolved and Had Simpler Sense/Antisense-encoded Ancestors*—By the same token, the unexpectedly sophisticated Urzyme catalytic proficiencies nevertheless reduce the largest kinetic barrier to protein synthesis  $\sim 10^5$ -fold more than necessary to assemble peptides from activated monomers in the absence of ribosomes (Fig. 4A). We conclude that the Urzymes represent a more advanced catalytic proficiency and evolutionary development than the earliest peptides catalyzing aminoacylation. Extended prior co-evolution with ancestral tRNAs implies, in turn, that both Class I and II Urzyme-like enzymes descended from still simpler, less proficient polypeptide catalysts that may have faced a stronger selective pressure to accelerate amino acid activation, as proposed earlier for different reasons (20).

An evident modularity of both Urzymes strengthens the implication that simpler catalysts preceded their emergence. The strongest published evidence for modularity is the widespread conservation of a nonpolar packing motif in the first  $\beta$ - $\alpha$ - $\beta$  crossover of  $>120$  Rossmannoid superfamily members (22), together with the critical role played by this motif in tryptophan activation by TrpRS (15). We explicitly used this modularity to identify strong vestigial traces of sense/antisense coding of Class I and II Urzymes (23). It therefore is probably more than coincidental that the Class I ATP binding site containing this motif at the C terminus and a modified P-loop (PXXXXHIGH) at the N terminus of the helix aligns precisely antisense to the Class II aaRS Motif 2 ATP binding site (10).

These observations afford strong circumstantial evidence that genetically coded, ATP-mobilizing catalysts only  $\sim 46$ -res-

idues long preceded both Urzymes. Substantial evidence that these earlier catalysts were also encoded by opposite strands of the same gene (23) constitutes an important leitmotif connecting these antecedents through the Urzymes to the contemporary enzymes (Fig. 6). Specifically, the evidence for sense/antisense complementarity implies that ancestral Class I and II aaRSs emerged simultaneously, not sequentially. That evidence reinforces the case that Urzyme-like molecules were actually on the evolutionary path to modern protein synthesis.

*Class I and II Urzyme Capabilities May Have Been Sufficient to Initiate Natural Selection of Globular Proteins*—Although Class I and II aaRSs differ markedly in primary, secondary, and tertiary structure (24), both Urzymes accelerate activation and acylation in the proportions observed for full-length enzymes (Fig. 5B). Remarkably, none of the variance in TS stabilization free energies can be attributed to whether the catalyst is derived from TrpRS or from HisRS. By this test, Class I and Class II Urzymes have statistically indistinguishable catalytic profiles for cognate amino acids. This parity seems especially significant in view of the underlying preference (10) of Class I aaRSs for large aliphatic amino acids (Leu, Val, Ile, and Met) and that of Class II aaRS for smaller, hydrophilic amino acids (Ala, Gly, Ser, Thr, His, and Pro). The chief difference between Class I and II appears to be that they process complementary (*i.e.* nonpolar *versus* polar) amino acid types.

The contrasting amino acid specificities of Class I and II aaRSs for nonpolar and polar amino acid substrates (10) lend additional significance to their parallel emergence and subsequent evolution. High Urzyme ATP and cognate tRNA affinities combined with weaker amino acid affinities appear suited to insert amino acids with appropriate water solubility at positions specified by a succession of increasingly specific codes (25), generating diverse, nonrandom, and yet functional molten globular translation products.

## Aminoacylating Urzymes Challenge the RNA World

These specificities suggest that the first operational RNA code (8) may have been as simple as two amino acid types, activated by a single pair of Class I and II aaRSs. Such a code may therefore have sufficed to initiate synthesis of nonrandom peptides with the potential to form molten globules (26–28) with a variety of secondary structures, depending on the binary pattern of polar *versus* nonpolar side chains specified by the code. Finally, such an evolving translation system could thus have continuously participated in correlated natural selection of improved proteins and more specific codes, without disrupting the integrity of existing proteins.

*Urzyme Catalytic Proficiencies, Modularities, and Sense/Antisense Genetic Coding Imply That Translation Began in a Peptide•RNA World*—According to the RNA World Hypothesis (1) RNA catalysts alone launched coded peptide synthesis, eventually producing superior protein catalysts that replaced RNA in virtually every sphere of biochemistry. Two considerations weaken the argument that decoding and protein synthesis arose uniquely from RNA.

First, whereas the contemporary ribosomal catalyst for peptidyl transfer is RNA, there is no evidence in extant biology for ribozyme-catalyzed nucleic acid synthesis, amino acid activation, or tRNA acylation. Thus, the hypothesis lacks clear ancestral lineages for most of the necessary components, relying instead on hypothetical ribozymal catalysts for which there is scant phylogenetic support. Yarus (2) and others adduced considerable apparent support for the hypothesis using SELEX exponential enrichment (29, 30) to produce contemporary surrogates for these hypothetical ribozymes. Specifically, RNA aptamers have been isolated that accelerate amino acid activation (31) and RNA aminoacylation (32), promote specific amino acid recognition (33), and peptide synthesis (34). However, selecting these aptamers required synthetic technology (*e.g.* the polymerase chain reaction) that was absent when life began.

A second, more serious weakness is that the ribozymal path to complexity depends critically on the early emergence of accurate replication, tying complexity tightly to fidelity. Such a path offers no mechanism to transcend what Koonin (3) calls “Eigen’s Cliff,” which arises from the need for nearly all progeny to closely resemble their parents to sustain function. For the RNA World, this is essentially impossible to envision. Whatever mechanisms polymerized nucleotides, elaborating the requisite ~1800 nucleotides of RNA thought necessary for accurate replication and protein synthesis has been estimated to be prohibitively slow, even requiring multiple universes (3). Mechanisms involving complementary, less specific peptide catalysts would doubtless have accelerated the process and represent more probable paths to biology.

The scenario that peptide catalysts and RNAs co-evolved from simpler precursors addresses both weaknesses. (i) We emulated ancestral gene reconstruction (35) to produce the Urzymes. Multiple three-dimensional structure alignments of superfamily members revealed earlier phylogenetic relationships than are accessible from multiple sequence alignments (Fig. 1); protein design stabilized and solubilized invariant core peptides as putative ancestral forms; and experimental catalytic activities are quantitative metrics for the likelihood of ancestry.

Ancestral sequence reconstructions demonstrating sense/anti-sense ancestry of Class I and II Urzymes (23) further reinforce the three-dimensional structural homologies (Fig. 1 and Ref. 18), affording strong phylogenetic evidence that the high Urzyme catalytic proficiencies and functional sufficiency imply that close relatives of these Urzymes participated in the early evolution of translation. (ii) We postulate that catalysis and coding co-evolved from an initial state with minimal information via the simultaneous growth of stereochemically complementary peptides and RNAs. Antiparallel polypeptides and polynucleotides adopt complementary helical conformations (36, 37), suggesting rudimentary catalytic templating mechanisms for mutual assembly of both polymers from resources thought to be present on primitive earth. The striking reciprocity of proteins and RNA in biology is consistent with our proposal: proteins exclusively catalyze nucleic acid synthesis; RNA catalyzes protein synthesis; and genetic messages are interpreted by the small ribosomal subunit (38), a ribonucleoprotein.

Our proposal avoids the Eigen Cliff because its most important restriction is not sequence fidelity but appropriate stereochemistry and relative chirality. Models (36) suggest that hydrogen bonding between L peptide carbonyl oxygen atoms and D ribose 2′OH groups (or the opposite relative configuration) can orient both the 3′-OH group for polymerization, accounting for 5′-3′ linkages in polynucleotides, and C-terminal peptide carboxyl oxygens for nucleophilic attack on activated amino acids. We suggest that structural and catalytic repertoires of short polypeptide hairpins were likely critical to eventually establishing the code, and perhaps even to the emergence of RNA itself (36).

Reciprocal templating of peptide and RNA double helices by stereochemical complementarity entails a natural form of sense/antisense genetic coding. As each RNA strand would have influenced the amino acid sequence primarily of one of the two peptide strands, both strands would have been templates for functional peptides, and vice versa. Complementarity enforces non-random but distinct sequences for peptides assembled by opposite RNA strands.

Under this alternative hypothesis, the earliest “polymerases” were naturally peptides, and the earliest ribosomes were short RNA oligonucleotides. Sense/antisense coding therefore connects this simpler framework to the realm of phylogenetically meaningful and experimentally testable genetic constructs described here. It is consistent with a peptide-accelerated and hence more probable scenario than the exclusive RNA world for the origins of codon-directed translation (Fig. 5). Thus, rather than an RNA world, life more likely began in a Peptide•RNA World (36, 38, 39), in which evolutionary development of catalysis and coding were intimately coupled via stereochemical thermodynamics.

---

*Acknowledgments*—We thank J. Perona for advice on tRNA substrate preparation and assays; E. First for helping to clarify the coding implications of Urzyme-catalyzed aminoacylation; and H. Fried, J. Perona, and M. Edgell for generously helping with the manuscript.

---

## REFERENCES

- Gilbert, W. (1986) *Nature* **319**, 618
- Yarus, M. (2011) *Life from an RNA World: The Ancestor Within*, pp. 132–155, Harvard University Press, Cambridge, MA
- Koonin, E. V. (2012) *The Logic of Chance: The Nature and Origin of Biological Evolution*, pp. 351–395, Pearson Education, FT Press Science, Upper Saddle River, NJ
- 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
- Schwob, E., and Söll, D. (1993) Selection of a “minimal” glutamyl-tRNA synthetase and the evolution of class I synthetases. *EMBO J.* **12**, 5201–5208
- Francklyn, C., and Schimmel, P. (1989) Aminoacylation of RNA mini-helices with alanine. *Nature* **337**, 478–481
- Francklyn, C., Musier-Forsyth, K., and Schimmel, P. (1992) Small RNA helices as substrates for aminoacylation and their relationship to charging of transfer RNAs. *Eur. J. Biochem.* **206**, 315–321
- Schimmel, P., Giegé, R., Moras, D., and Yokoyama, S. (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8763–8768
- 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
- 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
- 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
- Schroeder, G. K., and Wolfenden, R. (2007) The rate enhancement produced by the ribosome: an improved model. *Biochemistry* **46**, 4037–4044
- Wolfson, A. D., Pleiss, J. A., and Uhlenbeck, O. C. (1998) A new assay for tRNA aminoacylation kinetics. *RNA* **4**, 1019–1023
- SAS (2007) Version 6, SAS Institute, Cary NC
- 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
- Kirby, A. J., and Younas, M. (1970) The reactivity of phosphate esters: reactions of diesters with nucleophiles. *J. Chem. Soc. B* 1165–1172
- Schall, O. F., Suzuki, I., Murray, C. L., Gordon, J. L., and Gokel, G. W. (1998) Characterization of acyl adenyl anhydrides: Differences in the hydrolytic rates of fatty acyl-AMP and aminoacyl-AMP derivatives. *J. Org. Chem.* **63**, 8661–8667
- O'Donoghue, P., and Luthey-Schulten, Z. (2003) On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol. Mol. Biol. Rev.* **67**, 550–573
- Fournier, G. P., Andam, C. P., Alm, E. J., and Gogarten, J. P. (2011) Molecular evolution of aminoacyl tRNA synthetase proteins in the early history of life. *Orig. Life Evol. Biosph.* **41**, 621–632
- Härtlein, 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
- Xin, Y., Li, W., and First, E. A. (2000) Stabilization of the transition state for the transfer of tyrosine to tRNA(Tyr) by tyrosyl-tRNA synthetase. *J. Mol. Biol.* **303**, 299–310
- Cammer, S., and Carter, C. W., Jr. (2010) Six Rossmannoid 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
- Chandrasekaran, 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
- Carter, C. W., Jr. (1993) Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. *Annu. Rev. Biochem.* **62**, 715–748
- Delarue, M. (2007) KISS: the kinetoplastid RNA editing sequence search tool. *RNA* **13**, 161–169
- Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**, 1680–1685
- Moffet, D. A., Foley, J., and Hecht, M. H. (2003) Midpoint reduction potentials and heme binding stoichiometries of *de novo* proteins from designed combinatorial libraries. *Biophys. Chem.* **105**, 231–239
- Patel, S. C., Bradley, L. H., Jinadasa, S. P., and Hecht, M. H. (2009) Cofactor binding and enzymatic activity in an unevolved superfamily of *de novo* designed 4-helix bundle proteins. *Protein Sci.* **18**, 1388–1400
- Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510
- Ellington, A. D., and Szostak, J. W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822
- Kumar, R. K., and Yarus, M. (2001) RNA-catalyzed amino acid activation. *Biochemistry* **40**, 6998–7004
- Illangasekare, M., and Yarus, M. (1999) A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. *RNA* **5**, 1482–1489
- Yarus, M. (2000) RNA-ligand chemistry: a testable source for the genetic code. *RNA* **6**, 475–484
- Zhang, B., and Cech, T. R. (1997) Peptide bond formation by *in vitro* selected ribozymes. *Nature* **390**, 96–100
- Thornton, J. W. (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375
- Carter, C. W., Jr., and Kraut, J. (1974) A proposed model for interaction of polypeptides with RNA. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 283–287
- Church, G. M., Sussman, J. L., and Kim, S. H. (1977) Secondary structural complementarity between DNA and proteins. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1458–1462
- Harish, A., and Caetano-Anollés, G. (2012) Ribosomal history reveals origins of modern protein synthesis. *PLoS ONE* **7**, e32776
- Cech, T. (2009) Crawling out of the RNA world. *Cell* **136**, 599–602
- 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
- Berezovsky, I. N., Grosberg, A. Y., and Trifonov, E. N. (2000) Closed loops of nearly standard size: common basic element of protein structure. *FEBS Lett.* **466**, 283–286
- Joseph, D. R., and Muench, K. (1971) Tryptophanyl transfer ribonucleic acid synthetase of *Escherichia coli*. I. Purification of the enzyme and of tryptophan transfer ribonucleic acid. *J. Biol. Chem.* **246**, 7602–7609