Molecular Cell
Article

A Minimal TrpRS Catalytic Domain Supports Sense/Antisense Ancestry of Class I and II Aminoacyl-tRNA Synthetases

Yen Pham,[1] Li Li,[1] Aram Kim,[1] Ozgun Erdogan,[1] Violetta Weinreb,[1] Glenn L. Butterfoss,[1] Brian Kuhlman,[1] and Charles W. Carter, Jr.[1,*]
[1] Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
*Correspondence: carter@med.unc.edu
DOI 10.1016/j.molcel.2007.02.010

SUMMARY

The emergence of polypeptide catalysts for amino acid activation, the slowest step in protein synthesis, poses a significant puzzle associated with the origin of biology. This problem is compounded as the 20 contemporary aminoacyl-tRNA synthetases belong to two quite distinct families. We describe here the use of protein design to show experimentally that a minimal class I aminoacyl-tRNA synthetase active site might have functioned in the distant past. We deleted the anticodon binding domain from tryptophanyl-tRNA synthetase and fused the discontinuous segments comprising its active site. The resulting 130 residue minimal catalytic domain activates tryptophan. This residual catalytic activity constitutes the first experimental evidence that the conserved class I signature sequences, HIGH and KMSKS, might have arisen in-frame, opposite motifs 2 and 1 from class II, as complementary sense and antisense strands of the same ancestral gene.

INTRODUCTION

Understanding how the 20 contemporary aminoacyl-tRNA synthetases (aaRSs) arose by natural selection poses a challenge for evolutionary biology. These enzymes activate amino acids and translate the genetic code by specific acylation of cognate tRNAs (Ibba et al., 2005). Inherently ten million times slower than peptide bond synthesis, amino acid activation is the most significant kinetic barrier associated with the origin of protein synthesis.

Ancestral aaRSs therefore represent the emergence of a daunting catalytic activity that became key to genetic decoding in protein synthesis. What were these primitive activating enzymes like? Contemporary aaRSs are sophisticated enzymes, ranging widely in molecular weight (~330–900 amino acids) and quaternary structure (monomers, dimers, and homo- and heterotetramers; Carter, 1993). The reduced likelihood of generating such complexity spontaneously, for example from an "RNA world," argues that the ancestral forms were simpler. Moreover, there were probably two different kinds of primordial activating enzymes, as there are two distinct aaRS superfamilies, classes I and II, each including aaRSs for 10 of the 20 canonical amino acids (Cusack et al., 1990; Eriani et al., 1990).

Differences between the two classes encompass primary, secondary, and tertiary structures and important aspects of mechanism (Carter, 1993). The dominant amino acid specificities of contemporary class I and class II aaRSs (large, aliphatic; median $\Delta G_{xfr,\ Water\text{-}cyclohexane} = -1.59$ Kcal/mol versus small, polar; median $\Delta G_{xfr,\ Water\text{-}cyclohexane} = 2.67$ Kcal/mol [Wolfenden, 1983]) are complementary and are both necessary to make globular proteins. Moreover, the most highly conserved amino acids defining the active sites of each class are themselves translated largely by the opposite class (Rodin and Ohno, 1995). The two aaRS classes therefore likely derived from distinct ancestral enzymes, whose respective origins nevertheless coincided approximately in time.

There is significant antisense complementarity between the consensus coding sequences for the class-defining peptide signatures in each superfamily (Rodin and Ohno, 1995). That complementarity led Rodin and Ohno to propose the unusual hypothesis that class I and II aaRSs arose opposite one another on the same ancestral gene, accounting for their simultaneous appearance and, by the eventual duplication and radiation of that gene, the symmetric class division (Figure 1A).

A second piece of circumstantial evidence for their hypothesis is that, contrary to the conventional wisdom that only the sense strand encodes proteins (Houen, 1999), the contemporary *A. klebsiana* GDH gene actually encodes a second, unrelated protein, in-frame, from the antisense strand. Induction of glutamate dehydrogenase in response to altered ammonia metabolism is accompanied by expression of an HSP70 from the opposite strand, which coimmunoprecipitates with the dehydrogenase (LéJohn et al., 1994). The catalytic centers of the two *A. klebsiana* proteins are structural homologs, respectively, of the class I and II aaRS active sites (Carter and Duax, 2002). This homology establishes that opposite

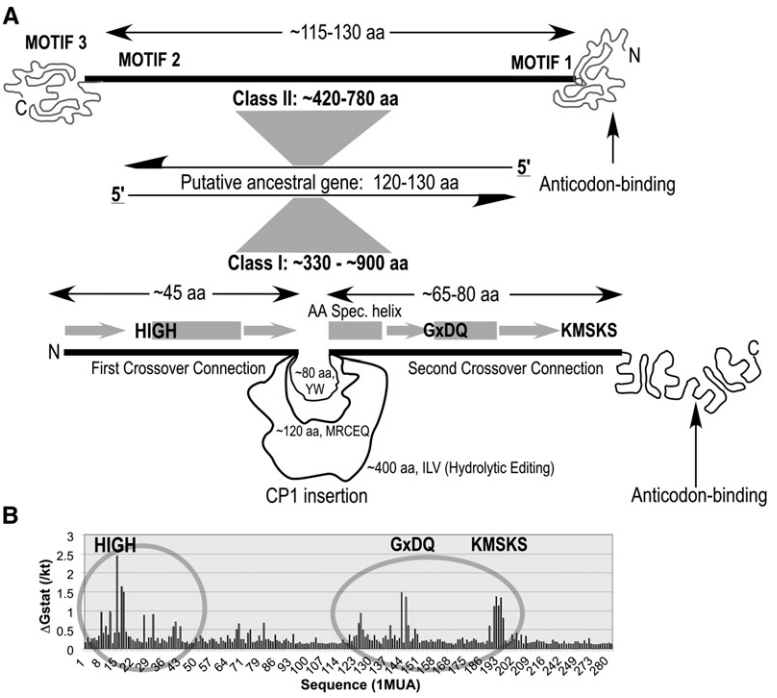

Figure 1. aaRS Domain Organization and Sequence Conservation and the Sense/Antisense Coding Hypothesis

(A) Class II core catalytic domains formed between motifs 1 and 2 are generally only ~100–130 amino acids long. Conserved class I catalytic sites (bold lines and gray secondary structures) are invariably interrupted by insertions, whose lengths vary from ~80 amino acids in TrpRS to ~400 amino acids in class Ia aaRSs specific for Ile, Leu, and Val and which are known as CP1 (Burbaum et al., 1990). Idiosyncratic anticodon binding domains occur at the C termini in class I and at either (but not both) termini in class II.

(B) Sequence conservation across class I aaRS MSAs. The ordinate, ΔG_{stat}, measures the log-likelihood associated with the frequencies of amino acids in each position of the MSA (Suel et al., 2003). High values indicate sequence conservation. The two gray ovals correspond to the two halves of the Rossmann fold (bold lines in [A]). Characters in boldface are active-site catalytic residues.

strands of the same gene can also encode folding instructions for the tertiary scaffolds required for proper spatial juxtaposition of the consensus, strand-complementary, active-site sequences.

Rodin and Ohno recognized that the sense/antisense-related signature sequences in contemporary aaRS genes are separated by quite different numbers of residues in the two classes and hence cannot be aligned opposite one another as suggested by their hypothesis (Figure 1A). The chief obstacle to constructing a sense/antisense gene is the presence of highly variable segments in class I aaRS known as "connective peptides 1 and 2" (CP1,2; Burbaum et al., 1990). They suggested that removing CP1 from the Rossmann fold would bring the consensus active-site signatures PxxxHIGH and KMSKS from class I much closer together, enabling them to align roughly opposite class II motifs 2 and 1, respectively. However, to date, this possibility has not been addressed experimentally.

Both aaRS classes possess modular structures that divide into conserved catalytic and idiosyncratic anticodon binding domains (Schimmel et al., 1993). Importantly, CP1 creates a further hierarchy by dividing all ten class I active sites into two discontinuous segments (Figure 1A). In some aaRS, system-specific editing modules are embedded within the catalytic domains; those in class I comprise a major component of CP1.

The central role of CP1 in class I aaRS gene differentiation encouraged us to examine multiple sequence and structural alignments to see if CP1 might be deleted entirely. We examined the mosaic structures within class I catalytic domains to identify an apparently minimal, ~130 amino acid, catalytic domain (MCD), derived by fusing discontinuous fragments from either side of CP1. This MCD is less than half the size of previously examined domain fragments (Augustine and Francklyn, 1997; Schwob and Söll, 1993). It is the largest structural subset that is conserved in all ten class I families, and it retains intact binding sites for both ATP and amino acid. We show here that it also retains substantial catalytic activity and describe additional properties of a possible sense/antisense double open reading frame resembling the putative ancestral synthetase gene.

RESULTS

Multiple Sequence and Structure Alignments Reveal a Conserved Minimal, Catalytic Domain

Class I aaRS sequences (1845 or ~185 for each aaRS family) from the HSSP database (http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+srsq2+-noSession) (Sander and Schneider, 1991) were first adjusted manually to ensure correct alignment of all signature sequences (Figure 1A). Then, 24 class I X-ray crystal structures were superimposed (Tetrada [Roach et al., 2005]; ProSup [Lackner et al., 2000]), assembled into a multiple structure alignment, and then consulted to assemble an overall class I aaRS multiple sequence alignment (MSA). Figure 1B summarizes primary sequence conservation within this MSA, numbered according to the *B. stearothermophilus* tryptophanyl-tRNA synthetase, TrpRS, sequence. The ordinate is the statistical conservation measure, ΔG_{stat} (Lockless and Ranganathan, 1999). Class I sequence conservation is dominated by three clusters of catalytic residues, PXXXHIGH (TIGN in TrpRS), GxDQ, and KMSKS,

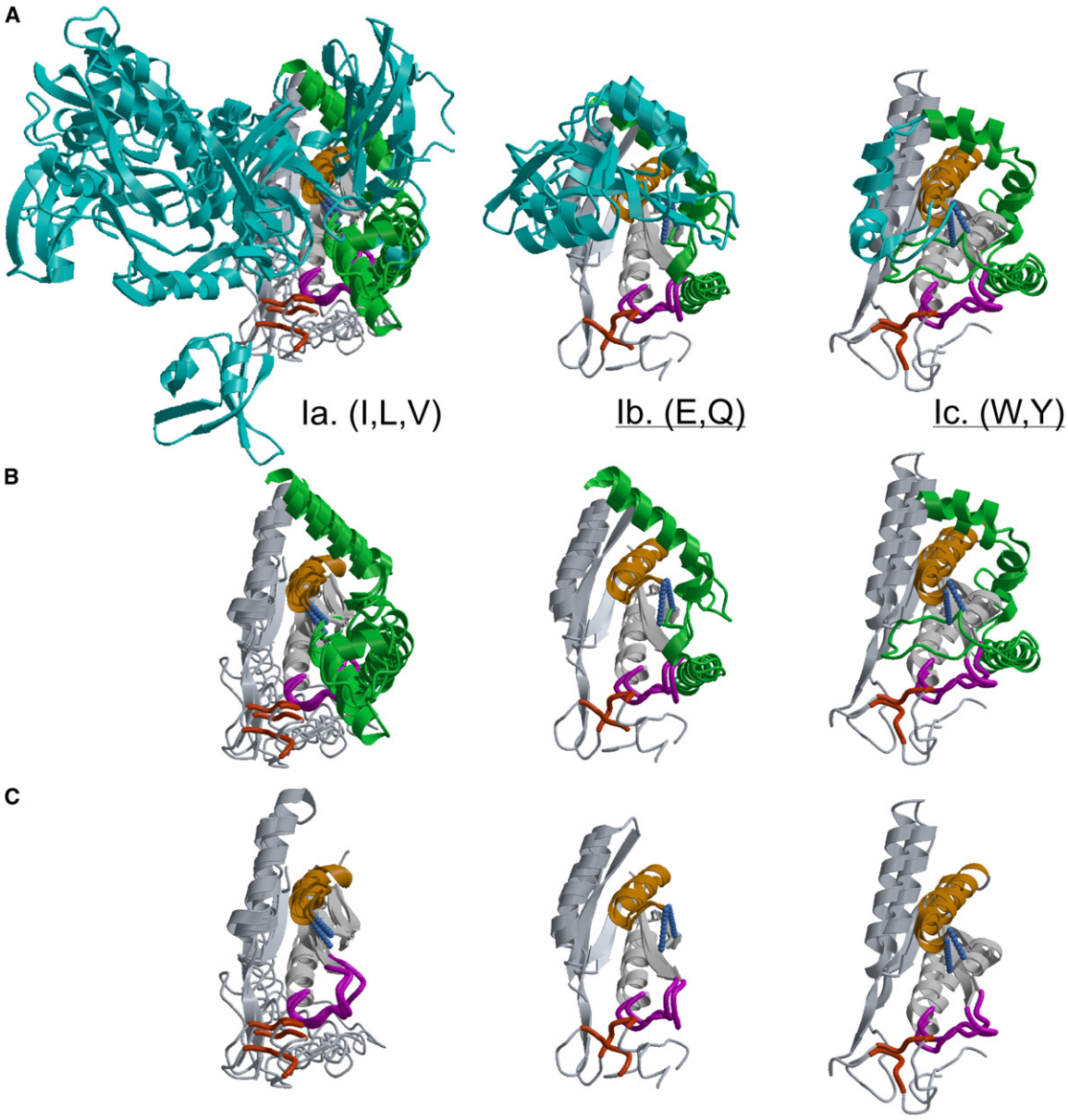

Figure 2. Hierarchical Mosaic Structure in Class I aaRS Catalytic Domain Tertiary Structures

(A) Catalytic domains of seven of the ten class I aaRS families, superimposed according to subgroups A, B, and C from left to right. Adaptive radiation (cyan) occurs almost exclusively in CP1.

(B) Portions of the CP1 peptide conserved in all class I aaRS (green).

(C) Highly conserved core domain consisting of the two crossover connections of the Rossmann fold (gray) with the specificity-determining α helix (orange), corresponding to the bold lines in Figure 1A.

and one cluster within the helix containing residues defining amino acid specificity (Praetorius-Ibba et al., 2000).

The sequence conservation in Figure 1B closely matches the structural conservation of both crossover connections and the specificity-determining α helix, which are common to all ten class I superfamily members, Figure 2C. The only significant structural homology across the CP1 elaborations in all subclasses IA, IB, and IC (Figure 2A) consists of three α helices, which are green/blue in Figures 2B and 3A. None of these helices appear to be essential to the active site.

Of special interest, the Cα atoms representing the beginning and end of CP1 are only ~5 Å apart in all ten class I aaRS. In TrpRS, CP1 (residues T46–G121) does not form a domain, sensu stricto. Rather, it wraps around the catalytic core as an exoskeleton or codomain (Figure 3). The conserved juxtaposition of CP1 termini in all class I aaRSs

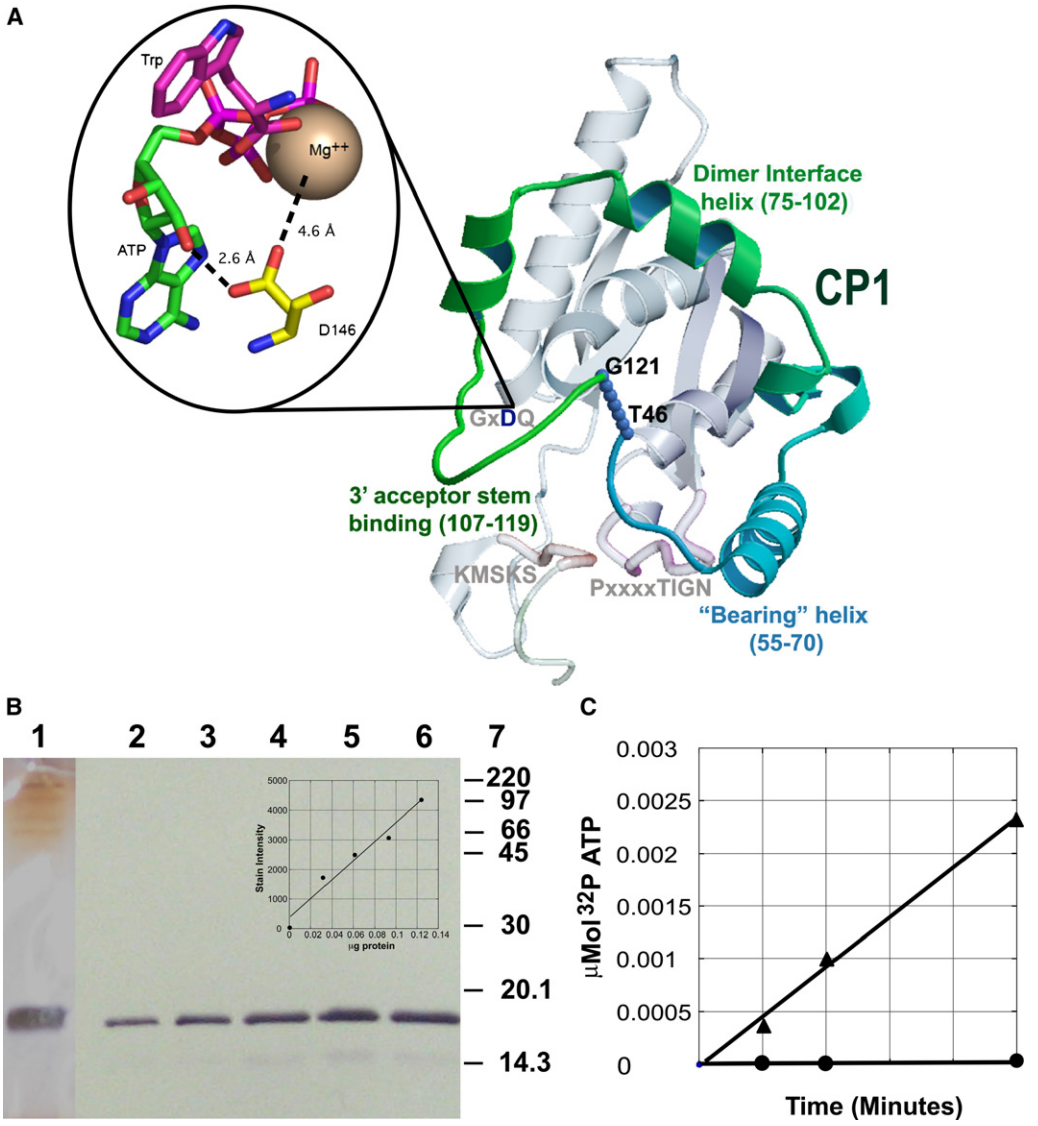

Figure 3. TrpRS Minimal Catalytic Domain

(A) Presumed structure, based on the 1MAU crystal structure, of the MCD (gray) and inserted exoskeletal, CP1, codomain. Functionalities associated with the TrpRS CP1 codomain (color) are indicated, together with their residue numbers, and are described in the discussion. Junctional residues where CP1 leaves and rejoins the MCD (black labels) are connected with a blue dotted line to represent the putative peptide bond fusing the gap. Consensus catalytic signatures are indicated in gray letters, except for D146, whose environment in crystal structures of intact TrpRS is highlighted in detail.

(B) Purification of the redesigned fragment after renaturation from the inclusion body fraction. Lane 1, silver-stained gel of the redesigned MCD; lanes 2–6, anti-FLAG immunoblot analysis of increasing MCD concentrations, used for the standard curve in the inset; lane 7, molecular weight standards, $\times 10^{-3}$.

(C) Time course of the redesigned fragment, ▲, and an empty-vector control, ●, purified by anti-FLAG antibody beads.

therefore suggests that CP1 might be deleted entirely and replaced by a single peptide bond.

Designed Genes for TrpRS MCDs

We designed two MCD genes derived from TrpRS, rather than a consensus sequence from all aaRS, in order to be able to assay tryptophan-dependent pyrophosphate exchange activity (Figure 3). Protein design techniques were key to constructing these genes. First, the free ends created by removing CP1 had to be fused with minimal disruption of the active site. To this end, we used the protein modeling program Rosetta to search the protein structure database for loop fragments matching the three or four residues on either side of the fusion (Rohl et al.,

2004). A number of matches were identified that did not require inserting any new amino acids. Next, the loop segments were refined in the context of the rest of the protein using a Rosetta protocol that iterates between backbone relaxation and sequence optimization (Kuhlman et al., 2003). The lowest-energy variant most similar to native TrpRS (WT) had five mutant residues (44–48; Table 1) and served as the template for the first synthetic gene.

Deleting the anticodon binding domain and CP1 codomain exposed numerous nonpolar side chains from TrpRS hydrophobic core regions. To compensate for this disruption, we again used Rosetta to evaluate additional mutations, excluding residues known to interact with Trp or ATP, to minimize deleterious effects of the deletion by (1) repairing hydrophobic cores surrounding the active site to increase stability and (2) introducing charged or polar side chains to enhance solubility. Twelve additional mutations resulted in a second, redesigned MCD gene (DES, Table 1). Notably, most designed mutations occurred at sites with low sequence conservation across the class I aaRS superfamily (Table 1 and Figure 1B).

Table 1. Mutations Introduced by Protein Design

Residue (WT Residue Number)	Wild-Type (WT)	Designed (DES)	ΔG_{stat}
13 (13)	Val	Met	0.13
14 (14)	Ile	Leu	0.40
16 (16)	Ile	Pro	0.42
19 (19)	Tyr	Trp	0.43
23 (23)	**Leu**	**Gly**	0.16
39 (30)	Ile	Met	0.31
44 (44)	*Ala*	*Tyr*	*0.18*
45 (45)	*Ile*	*Pro*	*0.15*
46 (46)	*Thr*	*Val*	*0.17*
47 (122)	***Leu***	***Glu***	*0.12*
48 (123)	***Leu***	***Thr***	*0.32*
61 (135)	Leu	Gly	0.16
79 (153)	**Leu**	**Lys**	0.19
87 (161)	**Phe**	**Tyr**	0.24
91 (165)	**Tyr**	**His**	0.92
94 (168)	Leu	Phe	0.13
130 (204)	Ile	Leu	0.41

Italicized residues (44–48) were common to both constructs and reflect remodeling of the turn created to fuse the N- and C-terminal fragments. The remaining residues represent changes introduced to compensate for removal of CP1 in the designed fragment. Boldface indicates changes expected to increase solubility. ΔG_{stat} is a measure of sequence conservation across the class I superfamily (Lockless and Ranganathan, 1999).

Minimal Catalytic Domains Derived from Bst TrpRS Are Catalytically Active

We implemented construction of the two TrpRS MCDs based on the schematic in Figure 3A by commercial gene syntheses (GenScript) as they involved a long internal deletion, and either five (WT) or 17 (DES) mutations, in addition to the C-terminal deletion. We added N-terminal FLAG tags for purification to avoid disrupting the C-terminal aliphatic residue, I/L204, which is embedded in a hydrophobic cluster that also includes I16 and M193 and links the TIGN and KMSKS signatures.

Cells transformed with plasmids encoding these genes rapidly lost the ability to express the cloned proteins, suggesting significant toxicity of the TrpRS fragments, perhaps resulting from loss of specificity and hence mischarging. Moreover, immunoblot analysis revealed a predominance of the expressed fragments in the insoluble fraction under a variety of different growth and induction conditions. Preliminary ^{32}PPi exchange tryptophan activation assays after purification on anti-FLAG antibody beads showed that the supernatant fractions were contaminated by host cell activity evident in an empty-vector control.

FLAG-tagged tryptophan activation activity could be renatured from 6 M urea extracts of the inclusion bodies from expression of the redesigned (DES) MCD fragment (Figure 3). Renaturation of the native TrpRS (WT) MCD produced instead a viscous, ill-behaved solution we have been unable to characterize. So we proceeded with the DES fragment, which was purified to near homogeneity from the inclusion body fraction (Figure 3B). A similar treatment of the empty-vector control had no detectable activity (Figure 3C). Quantitation of both silver-stained gels and immunoblots provided protein concentrations with an estimated error of 9.3% (Figure 3D). An incomplete factorial screen (Chester et al., 2004; Yin and Carter, 1996) identified improved renaturation conditions that increased specific activity 20-fold (Table 2).

To corroborate that the expressed MCD fragment is responsible for the catalytic activity observed in Figure 3C, we made the active-site mutant, D146A (Figure 3A, detail), which reduces the activity of full-length TrpRS by 200-fold. The corresponding TyrRS mutant is also among those with the largest impact on catalysis, and its effects are evident only in the transition state (Fersht, 1987). The specific activity of this mutant was actually ~20-fold higher than that of the native variant (Table 2). Although surprising, this change in specific activity validates the conclusion that the MCD fragment is the active catalyst. The unexpected activation of this mutant is also consistent with the observation that expression of this mutant MCD was greatly reduced, especially if, as suggested, more active MCD fragments are toxic to cells in which they are expressed because of higher levels of mischarging.

The fragments have less than 20% of the mass of the active TrpRS dimer. Their specific activities are ~10^{-5} and 10^{-4} times that of the native, dimeric enzyme; thus,

Table 2. Relative Catalytic Efficiencies for Amino Acid Activation

Protein	Specific Activity (Prod/mol/s) Max; Average ± SD (Replicates)	Relative Specific Activity
Wild-type TrpRS	64888	1
D146A full-length	326, 308 ± 25 (2)	0.005
Des MCD	1.36; 1.26 ± 0.16 (4)	0.00002
D146A Des MCD	26.5; 20.2 ± 8.9 (2)	0.0004
	k_{cat}/Km (Prod/mol/s)	
KK13 Ribozyme (Kumar and Yarus, 2001)	0.38	0.000006
	k_{non} (Prod/mol/s)	
Uncatalyzed rate[a]	8.3×10^{-9}	1.3×10^{-13}

Values for the MCD fragment and its D146A mutant were obtained from an incomplete factorial experiment (details not shown). Assays from the design that were close to the maximum value were averaged. Standard deviations are for experiments done under different renaturation conditions and therefore overestimate the variance of the resulting average. The maximum values were used to compute relative specific activities in the third column.

[a] An appropriate reference is provided by studies of catalytic rate enhancements, given by the ratio of k_{cat}/Km to the uncatalyzed rates k_{non} (Radzicka and Wolfenden, 1995). Amino acid activation is similar to the reaction of acetate with methyl-p-NO_2-phenyl phosphate, the p-NO_2-phenolate leaving group behaving similarly to the pyrophosphate leaving group in amino acid activation. The uncatalyzed rate of this reaction provides a close analog for amino acid activation. The published value for this rate is 7×10^{-6}/mol/min for reaction at pH 9.7, 39°C (Kirby and Younas, 1970). Correcting to temperature (25°C) and pH (7.5) gives a value of 5×10^{-7}/mol/min or 8.3×10^{-9}/mol/s (D. Herschlag and M. Forconi, personal communication). The apparent second-order rate constant for the aaRS ($\sim 10^6$/mol/s) corresponds to a rate acceleration of roughly 10^{14}. By comparison, the ribosome is a quite modest catalyst, accelerating peptidyl transfer by only about seven orders of magnitude over the uncatalyzed rate (Sievers et al., 2004).

they catalyze tryptophan activation with k_{cat}/Km values of 10^9 and 10^{10} times the uncatalyzed rate. They therefore retain substantial catalytic activity.

DISCUSSION

We were motivated by two problems associated with the origin of protein synthesis: finding simpler peptide catalysts of amino acid activation and examining the compatibility of minimal class I and II catalytic domains with simultaneous coding on opposite strands of one gene. The catalytic activity of the TrpRS MCDs is an important proof of principle in both respects. Further, this work outlines a sensible path from a simpler genetic code toward contemporary complexity and provides a tool kit and metric for exploring early enzyme evolution.

Small Polypeptides Can Exceed the Catalytic Proficiency of Engineered Ribozymes

Amino acid activation is a complex, bisubstrate reaction. AaRS catalytic domains of 320 residues in class II HisRS (74% of 434 aa; Augustine and Francklyn, 1997) and 260 residues in class I GlnRS (48% of 547 aa; Schwob and Söll, 1993) retain activity for both amino acid activation and aminoacylation of tRNA. Such constructs represent nearly intact contemporary domains and therefore seem more sophisticated than the putative ancestral catalysts. We demonstrate here that polypeptides half as long (~130 residues) could have activated amino acids substantially faster than the uncatalyzed rate. Notably, the D146A mutant is less than half the mass of, and also 70 times faster than, an extensively engineered, 114 nucleotide amino acid activating ribozyme (Table 2, KK13; k_{cat}/Km = 0.38/mol/s; Kumar and Yarus, 2001). This work therefore projects polypeptide-based catalysis considerably further back into the realm where ribozymes are currently thought to have functioned exclusively.

TrpRS MCD Catalytic Activity Verifies a Key Prediction of the Rodin-Ohno Hypothesis

Rodin and Ohno (1995) proposed that the two aaRS classes descended from opposite strands of the same sense/antisense double open reading frame (SASDORF; Duax et al., 2005). Although few examples of such highly efficient coding have been identified in the contemporary proteome, this unusual hypothesis nevertheless presents a logical paradigm with experimentally testable implications: it predicts (1) the existence of minimal, structurally viable and catalytically active homologs of the class I and class II catalytic domains, in which (2) the two sets of class-defining peptide signatures can be spaced appropriately on complementary strands in the primitive synthetase gene, consistent with being encoded by a SASDORF. The MCD catalytic activities in Table 2 verify both predictions, complementing the surprising evidence (Carter and Duax, 2002) that each of two contemporary proteins encoded by a verified SASDORF is structurally homologous to one of the two aaRS classes and therefore strengthening the sense/antisense hypothesis substantially.

Deleting CP1 and fusing the two discontinuous halves of the Rossmann fold substantially reduces the discrepancy between the separations of HIGH and KMSKS class I signatures in class I genes and class II motifs 1 and 2 in class II genes. Figure 4A illustrates an alignment of both class-defining peptide signatures opposite one another in the absence of CP1, as envisioned (Rodin and Ohno, 1995). This construction was generated from the actual gene sequences of *B. stearothermophilus* TrpRS and *E. coli* HisRS, using the sense coding strands of the TrpRS

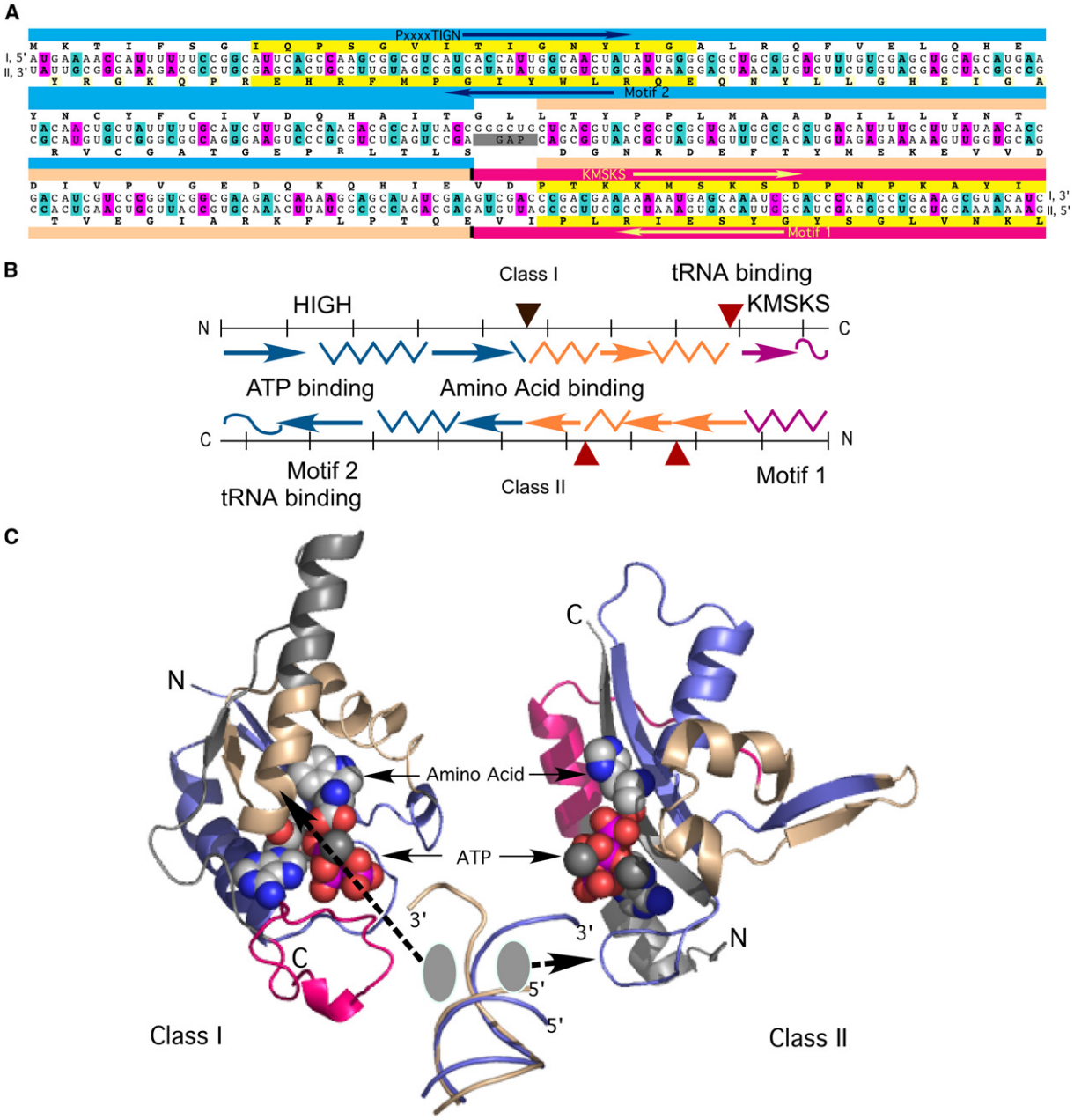

Figure 4. Illustrative Ancestral Sense/Antisense "Gene" for Minimal Catalytic Domains Coding that for Class I TrpRS Opposite that for Class II HisRS

(A) Nucleotide sequences for the catalytic domains drawn from the genes for *B. stearothermophilus* TrpRS (top strand) and *E. coli* HisRS (bottom strand) were aligned using GAP (Accelrys, 2006). The alignment matched both pairs of signature sequences to each other (yellow color and arrows), as expected from the Rodin-Ohno hypothesis. Colored bands above the sequences indicate three continuous sections of the HisRS gene that align antisense to the TrpRS gene sequences.

(B) Secondary structures encoded by the 96 amino acids in (A), including functionally important residues for ATP, amino acid, and tRNA binding. The location of the 76 residue CP1 peptide excised from TrpRS is indicated by the black inverted triangle midway along the gene. The red inverted triangle denotes the location of the 30 residue gap remaining in the HisRS sequence. Red triangles under the class II coding strand indicate locations in which the class II active sites have the most variable numbers of residues.

(C) Structures of the TrpRS and HisRS MCDs, colored as in (A) and (B), are based on the respective crystal structures and superimposed using the β strand loop regions culminating in the TIGN and motif 2 signatures, resulting in antiparallel orientations of the two active sites, the TrpRS MCD running roughly N to C from top to bottom and HisRS running from bottom to top. Gray segments are not encoded in (A). A superimposed pair of class I and II tRNA acceptor stems is indicated, with dashed arrows to primary binding surfaces in the two classes. As noted by Ribas de Pouplana and Schimmel (2001b), the tRNA binding sites approach from opposite directions (upper left in TrpRS, lower right in HisRS).

MCD (130 amino acids) and the sequence from motif 1 through motif 2 (96 amino acids) from the HisRS active site. As both the first and third bases of each codon entail a wobble base on one or the other strand, only the central base was retained from each codon. Alignment of the resulting sequences was carried out using the GAP function from the Accelrys GCG suite (Accelrys, 2006).

The alignment places 46 N-terminal amino acids of TrpRS (containing PxxxxTIGN) opposite the corresponding C-terminal 46 amino acids (containing motif 2) of HisRS (slate) and the 18 residue C- and N-terminal segments (KMSKS and motif 1) directly opposite one another (magenta). An intermediate 30 residue HisRS segment (wheat) closely precedes the C-terminal segment, leaving an ~30 residue gap following motif 1. This gap occurs precisely where the class II active sites have the greatest length heterogeneity, the corresponding segment ranging from five residues (1J5W; *T. maritima* GlyRS) to 120 residues (1ATI; *T. thermophilus* GlyRS). This range therefore affords ample opportunity for a structural match between the two classes.

All three segments had ~45% Watson-Crick pairing of middle bases, compared to an expected value of 25%. Assessing the significance of this value is nontrivial, among other reasons because gap positions are restricted by the sense/antisense relations identified earlier by Rodin and Ohno (1995) and because of nonrandom interactions between bases within a strand. Each possible registration of the intermediate fragment was analyzed. The final position was defined by a maximum of 14/25 base pairs for middle bases, 2.5 standard deviations greater than the average value of 7.6 base pairs over all possible positions. The paired middle base frequency of 0.45 appears sufficiently improbable to be noteworthy. A simulation of random alignments of coding frame hexamers, derived from the two strands to represent interstrand base interactions from both strands of the full-length gene, had an average middle-base complementarity of 0.27 ± 0.044, four standard deviations less than that in Figure 4A ($p = 0.0065$).

Both strands encode the primary surfaces (central α helix in the second crossover in class I and motif 2 loop in class 2) used by contemporary aaRS to bind tRNA acceptor stems. Figures 4B and 4C show schematically that the alignment organizes ATP and amino acid binding consistently with two notable characteristics of class-specific tRNA binding (Ruff et al., 1991). tRNA acceptor stem binding sites both occur at the respective C termini and are therefore parallel. On the other hand, sites for the two low molecular weight substrates occur in reverse order along the two strands and, moreover, lie opposite one another.

Thus, activated amino acids in the two classes are related by reflection across the "gene," whereas the acceptor stem binding sites are related by rotation of the two coding strands (Figure 4B). A curious structural correlate of this contrasting arrangement appears if the ATP binding sites are superimposed using the β strand loop regions culminating in the TIGN and motif 2 signatures (rmsd = 2.5 Å for ten Cα atoms). The activated amino acids aligned in this way are accessible only from opposite directions (Figure 4C), consistent with the opposite ribose hydroxyl group preferences of the two classes (Cusack, 1995; Cusack et al., 2002; Eriani et al., 1990). Moreover, the tRNA 3′ acceptor stem binding surfaces arise from behind and outside of the respective nucleotides. Thus, even this minimal configuration retains the potential for stereochemically compatible docking to opposite sides of cognate tRNA acceptor stems considered by Ribas de Pouplana and Schimmel (2001b).

Unexpected D146A Active-Site Mutational Effects Suggest a Change in Mechanism

Enhancement of the TrpRS MCD activity in the D146A active-site mutant was not anticipated and awaits verification by active-site titration. We consider it unlikely, however, that renaturation of the poorly expressed D146A mutant is 100 times more efficient than that of the D146 variant. Among the possible explanations for this surprising result is that the catalytic mechanisms of the full-length, dimeric TrpRS and the MCD differ. Indirectly, that explanation would imply that the additional domains in contemporary TrpRS facilitated the exploitation of a new catalytic role for D146 that was not available to the MCD but became useful only after long-range interdomain interactions had enabled evolving class I aaRS to exploit it.

From a Simpler Genetic Code to Contemporary Complexity

The detailed matching of aaRS catalytic and anticodon binding domains to the 3′ acceptor/TΨC and D/anticodon stems of their cognate tRNAs has suggested to others that the two domains in each biopolymer arose independently and fused after first functioning and coevolving separately (Ribas de Pouplana and Schimmel, 2001a; Steer and Schimmel, 1999). Consistent with independent development of aaRS catalytic and anticodon binding domains, more than half of the contemporary aaRSs use an anticodon-independent, "operational" code to specifically aminoacylate minihelices derived from the 3′-acceptor/TΨC domains of their cognate tRNAs (Francklyn et al., 1992; Francklyn and Schimmel, 1989; Musier-Forsyth and Schimmel, 1999.). Recognition of this "operational" code by specific aaRS catalytic domains suggests an early ribonucleoprotein environment in which tRNA acceptor stems and synthetase catalytic domain precursors, perhaps like the TrpRS MCD, functioned without reference to the anticodon (Schimmel et al., 1993). As tRNAs are among the oldest surviving biological macromolecules (Maizels and Weiner, 1994; Weiner and Maizels, 1987), that "collaboration" would have served as an important bridge between a prebiotic world in which RNA molecules may have played a more dominant role in translation (Joyce, 2002; Noller, 2004) and contemporary biology.

The mosaic hierarchy in Figure 2 is likely related to the development of the genetic code. The ascending rows suggest an evolutionary path for the adaptive radiation

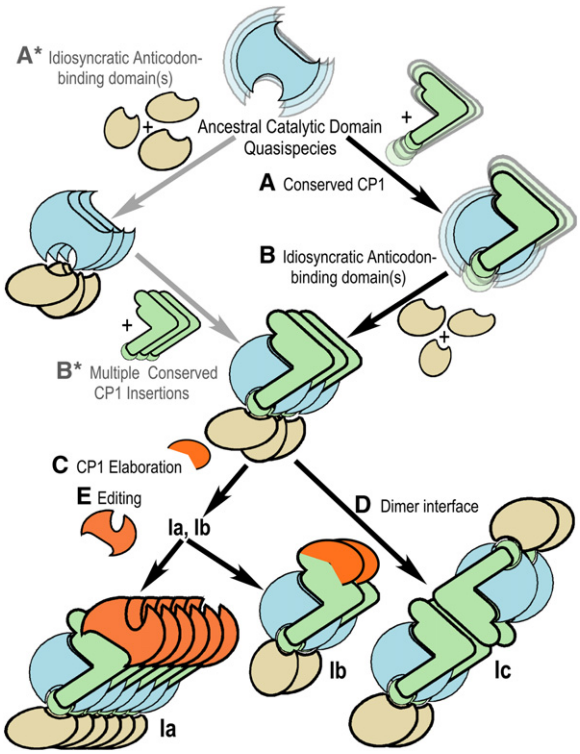

Figure 5. Alternative Early Evolutionary Paths for the Hierarchical Mosaic Structure of Class I aaRS

Colors, blue (ancient) → red (recent) encode relative age. The order in which the MCD assimilated the CP1 and anticodon binding domains is uncertain (A, B versus A*, B*). Conservation of the basic CP1 codomain argues, via maximum parsimony, that it preceded the idiosyncratic anticodon binding domains. Gray echos in the upper right-hand corner suggest quasispecies, or weighted average sequences of populations that may have cooperated with other quasispecies and were the targets of natural selection before the last common universal ancestor emerged (Eigen and Winkler-Oswatitch, 1992).

of class I aaRS, in which the intervening segments comprising CP1 were successive additions to a pre-existing MCD. The evident structural conservation argues strongly for divergence of the class I superfamily from a common ancestor, rather than convergence. Virtually all contemporary structural variation evident in class I aaRS Rossmann fold domains occurs within the CP1 insertion (Figures 1 and 2; O'Donoghue and Luthey-Schulten, 2003). The absence of the CP1 fragment from the MCD suggests that it may not have been required for the earliest class I aaRS. Moreover, as the TrpRS CP1 codomain is the smallest in the superfamily, and since its three helices occur in all class I aaRS, the TrpRS MCD and its intact catalytic domain may resemble distinct, early stages of class I superfamily evolution (Figure 5; right-hand branch).

The TrpRS CP1 codomain is highly amphipathic, contributing both internal hydrophobic core residues and charged residues on its surface. Thus, the primordial CP1 insertion probably stabilized and solubilized the rudimentary MCD. In TrpRS, it also serves three adaptive functions (Figure 3A). The final helix and loop, residues 107–119, are components of the acceptor binding domain of other class I enzymes, such as GlnRS. They have been implicated in catalysis in contemporary TrpRS (Kapustina and Carter, 2006) and TyrRS (Fersht, 1987) and likely therefore acquired functionality before the acquisition of anticodon binding domains. The N-terminal helix, residues 55–70, provides a bearing about which the TrpRS anticodon binding domain rotates in its conformational cycle, and it likely adapted only after that acquisition. The intermediate helix, residues 75–102, anchors the dimer interface. CP1 contributes roughly 60% of the TrpRS dimer interface, so the MCD probably functions as a monomer. As TrpRS and TyrRS from subclass Ic are likely the most recent aaRSs (Woese et al., 2000), this helix was probably the most recent functional specialization.

The divergent evolutionary model would therefore imply a period, either before or after acquisition of CP1, during which its specificity was restricted, perhaps as a quasispecies (Eigen and Winkler-Oswatitch, 1992), to a single type of amino acid, such as the aliphatic amino acids in the largest contemporary subclass, Ia. The important catalytic role of arginine in the active sites of all class II aaRS suggests that the earliest branching in class I may have involved differentiation of ArgRS from those specific for aliphatic side chains. These observations suggest that the ancestral class I aaRS speciated first into families for different class Ia amino acids. Contemporary subclass Ia amino acids are coded by an average of 3.7 codons per amino acid, compared to only two synonymous codons for subclasses Ib and Ic, consistent with early colonization of the genetic code by subclass Ia enzymes, followed first by Ib and then by Ic.

Implicitly, these early stages of class I aaRS evolution preceded the time when the genetic code reached its full complement of 20 canonical amino acids. The functionality of intermediate aaRSs may therefore have been achieved using a reduced amino acid alphabet (Trifonov, 2005). Two authors have recently discussed the stepwise growth of the genetic code (Delarue, 2007; Rodin and Rodin, 2006) that is implied by the functionality of a reduced alphabet.

Irrespective of the validity of the Rodin-Ohno hypothesis, the two parallel branches in Figure 5 raise a vexing question. The incompatibility of CP1 with sense/antisense coding of the two classes and the high concentration of side-chain specificity determinants within the MCD argue that gene duplication, giving rise in parallel to additional pairs of aaRS genes, took place before the insertion of CP1 (Figure 5, left branch). However, that conclusion is contradicted by the observation that the TrpRS CP1 codomain peptide is structurally homologous across the entire class I superfamily and, by maximum parsimony, was likely acquired before extensive gene doubling (Figure 5, right branch). Although several contemporary genetic mechanisms, including transposition and gene conversion, could have spread a common structural insertion across an established class I aaRS superfamily, it is not

our purpose here to consider what features of the class I domain architecture, if any, might have facilitated such mechanisms.

Concluding Remarks

The central importance of accelerating amino acid activation and the high conservation within the aaRS catalytic domain fragments we have fused to create the TrpRS MCD both argue that the present fragments resemble early ancestors of the class I aaRS. By efficiently harnessing the free energy of ATP pyrophosphorolysis to tryptophan activation, the TrpRS MCD has a key property of the protosynthetase catalytic domains proposed to function with tRNA 3′ acceptor/TΨC-like stem loops as the earliest "decoding pairs." Primordial catalysts resembling the TrpRS MCD would likely have accelerated protein synthesis significantly and had a substantial selective advantage over other catalysts present at that time, including putative ribozymes, and might have helped implement the "operational code" (Schimmel et al., 1993). Further, by providing transient protection of activated amino acids, this activity represents a sensible acylating agent for tRNA acceptor stem nucleophiles and would therefore drive eventual selection of increasingly specific mechanisms of tRNA recognition.

Sense/antisense coding of ancestral aaRS represents exceptionally tight genetic linkage between the genes on opposite strands. From a contemporary perspective, this linkage seems unduly restrictive and hence costly, highlighting questions about the possible selective advantage such efficient coding may have brought to prebiotic protein synthesis. In a precellular environment, however, such linkage would have ensured that both aaRS classes, and hence a sufficient repertoire of complementary amino acid types, were available simultaneously for translation and eventual natural selection of globular proteins. The complementary coding capacities in classes I and II, the detailed sense/antisense complementarity between codons for core and surface amino acids (Zull and Smith, 1990), and the high proportion of functional molten globules in random, binary-patterned combinatorial libraries (Kamtekar et al., 1993; Moffet et al., 2003) together suggest that such a gene might have sufficed to initiate natural selection.

Finally, the modules in Figure 5 provide a tool kit for reverse engineering contemporary class I aaRS, while the substantial gap between MCD and native TrpRS catalytic rates provides a metric for assessing successive selectable advantages associated with the accretion and mutational enhancement of the additional fragments in their mosaic hierarchy. The surprising D146A effects imply that the tool kit and metrics can also help reveal how newly acquired modules affected the catalytic mechanism. This work therefore establishes an experimental system for the study of quite early enzyme evolution, close to the limit of what can be studied experimentally with traditional genetic construction and biochemistry based on contemporary enzymes.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Plasmid Construction

Restriction enzymes and other molecular biology reagents were obtained from New England Biolabs and were used according to the manufacturer's protocols. Synthetic genes with native and redesigned sequences were synthesized commercially with 5′ NdeI and 3′ HindIII restriction sites and cloned into vector PUC57 by GenScript. An N-terminal FLAG tag was added by PCR. Following digestion with the appropriate enzymes, the PCR products were cloned to *E.coli* expression plasmid pET-42*a* for use in bacterial strain BL21(DE3)pLysS (Novagen). The D146A mutant in the DES fragment was prepared using the Invitrogen GeneTailor Site-Directed Mutagenesis System. All constructs were confirmed by DNA sequencing.

Protein Expression

All TrpRS fragments were expressed directly after transforming with plasmid DNA. After transformation, 100 μl competent cells were incubated in 1ml LB media at 37°C for 1 hr with shaking and then used to inoculate 100 ml growth media (LB plus 2% glucose, 30 mg/ml kanamycin, 34 mg/ml chloramphenicol). The inoculum was shaken at 250 rpm overnight at 37°C and then transferred to 1 l fresh LB with 30 mg/ml kanamycin and 34 mg/ml chloramphenicol. At OD_{600} = 0.4–1, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 1 mM final concentration to induce gene expression. Harvest began 14 hr after induction.

Purification of Fragments from Supernatant and Inclusion Bodies

A cell-free extract was prepared from frozen cell pellets by lysing cells in 50 mM Tris, 10% sucrose (pH 7.5), using a continuous flow pressure cell (Avestin), followed by centrifugation at 20,000 rpm for 30 min at 4°C, and the pellets were re-extracted in the same manner. Supernatants were combined and centrifuged at 27,500 rpm for 1 hr to yield the soluble fraction. The pellet (inclusion bodies) from this step and the combined supernatant were used as starting materials for purification. The soluble fraction was then dialyzed against anti-FLAG washing buffer (1× PBS, 0.05% Tween-20, 5 mM ATP, 50 μM tryptophan, 10% glycerol, 50 mM KCl, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 4 mM $MgCl_2$ [pH 7.8]).

Renaturation from Inclusion Bodies

Optimized renaturation conditions were determined from an incomplete factorial design (Chester et al., 2004; Yin and Carter, 1996) of nine different sets of conditions in which the temperature, protein concentration, KCl concentration, pH, dialysis protocol, and presence of substrates were varied randomly and rebalanced. Inclusion bodies were first dissolved and diluted in solubilization buffer (20 mM HEPES [pH 7.8], 6 M urea, 50 mM or 250 mM KCl, 5 mM $MgCl_2$, 1 mM EDTA, 1 mM PMSF, 1 mM β-mercaptoethanol, 5% glycerol [Campbell and Carpenter, 1995] with and without substrates) to OD_{280} = 2–10 (approximate total protein concentration of 2–10 mg/ml). Supernatants after further centrifugation (17,000 rpm for 30 min at 4°C) were dialyzed three times against ten volumes of the same buffer at 4° or 37°C, without urea. Before the final dialysis, bags were opened and 10 μl of anti-FLAG beads (Sigma) were added per ml of protein. FLAG-bound material was eluted in the same solution with 50 μg/ml of FLAG peptide solution. Multiple regression of specific activities against the experimental variables identified three factors (low protein concentration, high pH, and low KCl concentration) responsible for (R^2 = 0.87) of the variation in the data with t test probabilities of $\sim 10^{-4}$.

Purification on Anti-FLAG Resin Beads

Prewashed anti-FLAG bead suspension was added (10 μl/ml) to supernatants, which were shaken overnight at 4°C. Fragments bound to the beads (recovered in Bio-Rad Hi-Trap Q columns) were eluted with FLAG peptide (0.5 μl/ml).

Quantitation of Specific Activity

Both immunoblotting with anti-FLAG antibody and silver staining were used to construct standard curves for estimating protein concentration. As the silver-stained gels showed that the samples were quite pure, we determined most concentrations by using silver staining.

^{32}PPi Exchange Assays for Tryptophan Activation

Fragment and empty-vector eluates from anti-FLAG beads were concentrated 8- to 25-fold before being assayed. Samples were assayed using the traditional ^{32}P-pyrophosphate exchange as described (Joseph and Muench, 1971) except that a 3-fold reduction of background ^{32}P counts was achieved by collecting the charcoal containing labeled ATP on disposable spin columns, washing, and eluting the bound ATP with 50 μl of pyridine at 38°C.

ACKNOWLEDGMENTS

This work was supported by the National Institute of General Medical Sciences, 48519 and 78227. We are grateful to E. First and M.H. Edgell for continuing discussions throughout this work; to M.H. Edgell for suggesting the potential advantage of genetic linkage; to A. Danchin for advice on significance testing of the gene in Figure 4A; to R.F. Doolittle and M. Delarue for discussion; and to C. Francklyn, R. Wolfenden, R. Doolittle, and L. Kaguni for their comments on the manuscript.

Received: September 2, 2006
Revised: January 3, 2007
Accepted: February 5, 2007
Published: March 22, 2007

REFERENCES

Accelrys (computer software) (2006). GCG, Accelrys Inc.

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.

Burbaum, J.J., Starzyk, R.M., and Schimmel, P. (1990). Understanding structural relationships in proteins of unsolved three-dimensional structure. Proteins *7*, 99–111.

Campbell, S., and Carpenter, J.W. (1995). Refolding and purification of *Ras* proteins. Methods Enzymol. *255*, 3–13.

Carter, C.W., Jr. (1993). Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. Annu. Rev. Biochem. *62*, 715–748.

Carter, C.W., Jr., and Duax, W.L. (2002). Did tRNA synthetase classes arise on opposite strands of the same gene? Mol. Cell *10*, 705–708.

Chester, A., Weinreb, V., Carter, C.W., Jr., and Navaratnam, N. (2004). Optimization of apolipoprotein B mRNA editing by APOBEC1 apoenzyme and the role of its auxiliary factor, ACF. RNA *10*, 1399–1411.

Cusack, S. (1995). Eleven down and nine to go. Nat. Struct. Biol. *2*, 824–831.

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.

Cusack, S., Tukalo, M., and Yaremchuk, A. (2002). Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. EMBO J. *21*, 3829–3840.

Delarue, M. (2007). An asymmetric underlying rule in the assignment of codons: possible clue to a quick early evolution of the genetic code via successive binary choices. RNA *13*, 1–9.

Duax, W.L., Huether, R., Pletnev, V., Langs, D., Addlagatta, A., Connare, S., Habegger, L., and Gill, J. (2005). Rational genomics I: antisense open reading frames and codon bias in short chain oxido reductase enzymes and the evolution of the genetic code. Proteins *61*, 900–906.

Eigen, M., and Winkler-Oswatitch, R. (1992). Steps Towards Life: A Perspective on Evolution (Oxford: Oxford University Press).

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.

Fersht, A.R. (1987). Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. Biochemistry *26*, 8031–8037.

Francklyn, C., and Schimmel, P. (1989). Aminoacylation of RNA minihelices 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.

Houen, G. (1999). Evolution of the genetic code: the nonsense, antisense, and antinonsense codes make no sense. Biosystems *54*, 39–46.

Ibba, M., Francklyn, C., and Cusack, S. (2005). Aminoacyl-tRNA Synthetases (Georgetown, TX: Landes Biosciences).

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.

Joyce, G. (2002). The antiquity of RNA-based evolution. Nature *418*, 214–221.

Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M., and Hecht, M.H. (1993). Protein design by binary patterning of polar and non-polar amino acids. Science *262*, 1680–1685.

Kapustina, M., and Carter, C.W., Jr. (2006). Computational studies of tryptophanyl-tRNA synthetase ligand binding and conformational stability. J. Mol. Biol. *362*, 1159–1180.

Kirby, A.J., and Younas, M. (1970). Reactivity of phosphate esters—reactions of diesters with nucleophiles. J. Chem. Soc. B *6*, 1165.

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.

Kumar, R.K., and Yarus, M. (2001). RNA-catalyzed amino acid activation. Biochemistry *40*, 6998–7004.

Lackner, P., Koppensteiner, W.A., Sippl, M.J., and Domingues, F.S. (2000). ProSup: a refined tool for protein structure alignment. Protein Eng. *13*, 745–752.

LéJohn, H.B., Cameron, L.E., Yang, B., and Rennie, S.L. (1994). Molecular characterization of an NAD-specific glutamate dehydrogenase gene inducible by L-glutamine: antisense gene pair arrangement with L-glutamine-inducible heat shock 70-like protein gene. J. Biol. Chem. *269*, 4523–4531.

Lockless, S.W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. Science *286*, 295–299.

Maizels, N., and Weiner, A.M. (1994). Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. Proc. Natl. Acad. Sci. USA *91*, 6729–6734.

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.

Musier-Forsyth, K., and Schimmel, P. (1999). Atomic determinants for aminoacylation of RNA minihelices and relationship to genetic code. Acc. Chem. Res. *32*, 368–375.

Noller, H. (2004). The driving force for molecular evolution of translation. RNA *10*, 1833–1837.

O'Donoghue, P., and Luthey-Schulten, Z. (2003). On the evolution of structure in aminoacyl-tRNA synthetases. Microbiol. Mol. Biol. Rev. *67*, 550–573.

Praetorius-Ibba, M., Stange-Thomann, N., Kitabatake, M., Ali, K., Soll, I., Carter, C.W., Jr., Ibba, M., and Soll, D. (2000). Ancient adaptation of the active site of tryptophanyl-tRNA synthetase for tryptophan binding. Biochemistry *39*, 13136–13143.

Radzicka, A., and Wolfenden, R. (1995). A proficient enzyme. Science *267*, 90–93.

Ribas de Pouplana, L., and Schimmel, P. (2001a). Aminoacyl-tRNA synthetases: potential markers of genetic code development. Trends Biochem. Sci. *26*, 591–596.

Ribas de Pouplana, L., and Schimmel, P. (2001b). Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. Cell *104*, 191–193.

Roach, J.M., Sharma, S., Kapustina, M., and Carter, C.W., Jr. (2005). Structure alignment via Delaunay tetrahedralization. Proteins *60*, 66–81.

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.

Rodin, S.N., and Rodin, A. (2006). Partitioning of aminoacyl-tRNA synthetases in two classes could have been encoded in a strand-symmetric RNA world. DNA Cell Biol. *25*, 617–626.

Rohl, C., Strauss, C.E.M., Misura, K.M.S., and Baker, D. (2004). Protein structure prediction using Rosetta. Methods Enzymol. *383*, 66–93.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., and Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. Science *252*, 1682–1689.

Sander, C., and Schneider, R. (1991). Database of homology derived protein structures and the structural meaning of sequence alignment. Proteins *9*, 56–68.

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. USA *90*, 8763–8768.

Schwob, E., and Söll, D. (1993). Selection of a 'minimal' glutaminyl synthetase and the evolution of class I synthetases. EMBO J. *12*, 5201–5208.

Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R. (2004). The ribosome as an entropy trap. Proc. Natl. Acad. Sci. USA *101*, 7897–7901.

Steer, B.A., and Schimmel, P. (1999). Domain-domain communication in a miniature archaebacterial tRNA synthetase. Proc. Natl. Acad. Sci. USA *96*, 13644–13649.

Suel, G.M., Lockless, S.W., Wall, M.A., and Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat. Struct. Biol. *10*, 59–69.

Trifonov, E. (2005). Theory of early molecular evolution: predictions and confirmations. In Discovering Biomolecular Mechanisms with Computational Biology, F. Eisenhaber, ed. (Georgetown, PA: Landes Biosciences).

Weiner, A.M., and Maizels, N. (1987). tRNA-like structures tag the 3′ ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. Proc. Natl. Acad. Sci. USA *84*, 7383–7387.

Woese, C.R., Olsen, G.J., Ibba, M., and Sôll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. *64*, 202–236.

Wolfenden, R. (1983). Waterlogged molecules. Science *222*, 1087–1093.

Yin, Y., and Carter, C.W., Jr. (1996). Incomplete factorial and response surface methods in experimental design yield optimization of $tRNA^{Trp}$ from in vitro T7 RNA polymerase transcription. Nucleic Acids Res. *24*, 1279–1286.

Zull, J.E., and Smith, S.K. (1990). Is genetic code redundancy related to retention of structural information in both DNA strands? Trends Biochem. Sci. *15*, 257–261.