

Tryptophanyl-tRNA Synthetase Urzyme

A MODEL TO RECAPITULATE MOLECULAR EVOLUTION AND INVESTIGATE INTRAMOLECULAR COMPLEMENTATION*

Received for publication, April 22, 2010, and in revised form, September 14, 2010 Published, JBC Papers in Press, September 23, 2010, DOI 10.1074/jbc.M110.136911

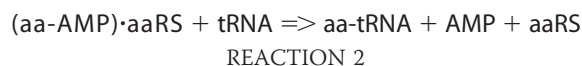
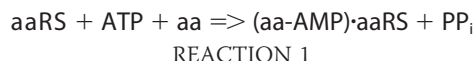
Yen Pham[‡], Brian Kuhlman[‡], Glenn L. Butterfoss[§], Hao Hu[¶], Violetta Weinreb[‡], and Charles W. Carter, Jr.^{‡1}

From the [‡]Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27599, the [§]Biology and Courant Computer Science Department, New York University, New York, New York 10003, and the [¶]Chong Yuet Ming Chemistry Building, University of Hong Kong, Pokfulam Road, 999077 Hong Kong, China

We substantiate our preliminary description of the class I tryptophanyl-tRNA synthetase minimal catalytic domain with details of its construction, structure, and steady-state kinetic parameters. Generating that active fragment involved deleting 65% of the contemporary enzyme, including the anticodon-binding domain and connecting peptide 1, CP1, a 74-residue internal segment from within the Rossmann fold. We used protein design (Rosetta), rather than phylogenetic sequence alignments, to identify mutations to compensate for the severe loss of modularity, thus restoring stability, as evidenced by renaturation described previously and by 70-ns molecular dynamics simulations. Sufficient solubility to enable biochemical studies was achieved by expressing the redesigned Urzyme as a maltose-binding protein fusion. Michaelis-Menten kinetic parameters from amino acid activation assays showed that, compared with the native full-length enzyme, TrpRS *Urzyme* binds ATP with similar affinity. This suggests that neither of the two deleted structural modules has a strong influence on ground-state ATP binding. However, tryptophan has 10³ lower affinity, and the *Urzyme* has comparably reduced specificity relative to the related amino acid, tyrosine. Molecular dynamics simulations revealed how CP1 may contribute significantly to cognate amino acid specificity. As class Ia editing domains are nested within the CP1, this finding suggests that this module enhanced amino acid specificity continuously, throughout their evolution. We call this type of reconstructed protein catalyst an *Urzyme* (Ur prefix indicates original, primitive, or earliest). It establishes a model for recapitulating very early steps in molecular evolution in which fitness may have been enhanced by accumulating entire modules, rather than by discrete amino acid sequence changes.

Aminoacyl-tRNA synthetases (aaRS)² include two superfamilies, each consisting of ~10 members. Both aaRS classes catalyze amino acid activation, the most difficult reaction in

protein synthesis, and thus are crucial to understanding the origin of the proteome. As suggested by their name, their chief function is to synthesize specific aminoacyl-tRNAs (Reaction 2), in effect translating the genetic code (2). In the absence of these biological catalysts, the first half of this reaction, amino acid activation by ATP (Reaction 1) occurs at a rate of ~8.3 × 10⁻⁹/mol/s (as corrected from the published value of 7 × 10⁻⁶/mol/min for reaction at pH 9.7, 39 °C for a similar reaction (3)). This represents, by many orders of magnitude, the highest activation energy barrier of any reaction required for protein synthesis. Hence, although the specificity of tRNA aminoacylation eventually became the hallmark activity of these enzymes (4), the emergence of amino acid-activating enzymes is probably the most significant requirement for launching the evolution of protein synthesis.



aaRSs very likely evolved from simpler polypeptide chains into complex multidomain, multifunction contemporary enzymes by progressively accumulating additional modular genetic information (5, 6), whose most obvious selective advantages arose from their ability to enhance catalytic activity and fidelity. Application of Ockham's razor suggests that vestiges of earlier forms should remain in contemporary enzymes. Structural analyses indicate that both class I and II aaRSs contain small (~120 amino acids) catalytic cores (2) complemented by internal insertions and external domains. We refer to constructs derived from these putative ancestral enzymatic cores as *Urzymes*, from Ur meaning primitive, earliest, or original plus enzyme. This term emphasizes that such constructs represent a ubiquitously conserved, dramatically smaller, and catalytically active polypeptide containing the active site of the corresponding native enzyme.

As *Urzymes* are identified by tertiary structural superposition, they should be distinguished from ancestral nodes resurrected from multiple sequence phylogenies (7–10). Genes resurrected in this manner represent earlier versions of what are essentially contemporary proteins. *Urzymes*, on the other hand, possibly represent the earliest protein catalysts.

One of the most striking successes of ancestral gene resurrection comes from work on the steroid hormone receptors (7,

* This work was supported, in whole or in part, by National Institutes of Health Grant GM078227.

¹ To whom correspondence should be addressed: 3092 Genetic Medicine, 120 Mason Farm Rd., Chapel Hill, NC 27599-7620. Fax: 919-966-2852; E-mail: carter@med.unc.edu.

² The abbreviations used are: aaRS, aminoacyl-tRNA synthetases; TrpRS, tryptophanyl-tRNA synthetase; CTD, C-terminal (anticodon binding) domain; CP1, connecting peptide 1; DES, redesigned TrpRS *Urzyme*; MD, molecular dynamics; MBP, maltose-binding protein; SUMO, small ubiquitin-related modifiers.

11). The authors (7, 11) used sequences derived by statistical inference from phylogenetic analysis of multiple sequence alignments to construct an ancestral gene. Then, by making multiple point mutations, Thornton and co-workers (11) modulated the receptor specificity for different ligands in keeping with the chronological divergence of ligand specificity along the phylogenetic tree. Finally, they built and tested a progressive set of mutations mimicking the evolution of the two descendant receptors, revealing that some of these made the pathways irreversible (12).

The TrpRS Urzyme construction was motivated by objectives shared with those of ancestral gene resurrection (8), namely to study the recapitulation of putative evolutionary events. In the case of the aaRS, such events include the development of amino acid specificity, the emergent ability to acylate tRNAs, and the subsequent development of cognate tRNA specificity. To mimic changes associated with these activities that might have occurred early along the evolution leading to contemporary enzymes, it is especially important that the effects of those changes also be measurable. This study documents the authenticity of such experimental measurements.

Of particular interest for this work is an inserted domain of variable length within the Rossmann dinucleotide binding fold in class I aaRS, called connecting peptide 1 (CP1) (1, 13). In TrpRS, the smallest class Ic aaRS, the CP1 insertion has only 74 residues and has no editing activity. However, data provided here point to the possibility that CP1 does indeed enhance amino acid specificity. Moreover, in three of the large class Ia aaRSs (LeuRS, IleRS, and ValRS), the corresponding internal insertion has grown into an independently structured domain harboring the editing function that is specific for the most frequently mistaken naturally occurring amino acids (14–17), thereby amplifying the fidelity of aminoacylation in those contemporary class I enzymes.

Simultaneous emergence of two classes of aaRSs has been rationalized by a unique hypothesis proposed by Rodin and Ohno (18). Based on multiple sequence alignments, they found that the codons for the HIGH and KMSKS class I signatures are very nearly antisense to the corresponding codons of the class II motifs 2 and 1, respectively. In other words, when the intervening CP1 insertions are removed, the KMSKS and HIGH loops can be approximately aligned, antiparallel with motif 1 and 2, suggesting that the class I and II core catalytic apparatuses could once have been gene products from the sense and antisense strands of one gene. The earliest aaRS would have been active enzymes; otherwise, they could not have participated in natural selection. The Rodin/Ohno observation therefore predicts that an appropriately constructed class I aaRS gene lacking CP1 and the C-terminal anticodon-binding domain (CTD) should be enzymatically active. However, until recently (19), no experiment had been done to test that prediction.

In our previous paper (19), we presented a preliminary description of such an experiment, in which we deleted both the C-terminal anticodon-binding domain and the CP1 peptide, fusing together the first and second halves of the TrpRS Rossmann fold. The resulting 130-residue fragment contained a nearly intact active site in which the TIGN and KMSKS signatures are separated by approximately the same number of

amino acids as class II motifs 2 and 1, respectively. Two constructs were expressed, one with a nearly native sequence and the other with 12 mutations from a re-design by Rosetta (20). Both were insoluble. However, the re-designed construct (DES), but not the native one (WT), could be renatured. Its specific activity represented at least a 10^9 -fold increase over the uncatalyzed rate, which is within 5 orders of magnitude of the native TrpRS activity (19).

This paper follows up that work (19) by describing important and previously unreported details of the design process, evidence regarding the DES TrpRS Urzyme structure from molecular dynamics (MD), and an expression system that produces that construct in soluble form. Having sufficient quantities of soluble DES-TrpRS, Urzyme has enabled us to titrate the fraction of active sites and determine Michaelis-Menten parameters, which differ markedly from those of the full-length enzyme, reinforcing the conclusion that the observed activity arises from the Urzyme itself.

EXPERIMENTAL PROCEDURES

Linker Design to Replace Connecting Peptide 1—The TrpRS Urzymes described in Pham *et al.* (19) were constructed *in silico* by fusing two halves of the Rossmann fold. Connecting peptide 1, a 74-residue-long fragment between Thr-46 and Gly-121, and the last 124 amino acids of the C-terminal domain were removed. The resulting 130-residue sequence was submitted to Rosetta design with six alternative linkers (A, AA, AGA, AAAA, AAGAA, and AGAAGA) between the disjoint active-site fragments. These linkers varied in length and consisted of short side chain amino acids alanine and glycine. A zero length linker was also included. For each linker, 60 backbone templates were identified to restore molecular continuity and designed to optimize the energy scores. To preserve the wild type sequence as far as possible (WT Urzyme), mutations were allowed in up to 4 residues adjacent to each side of CP1 removal.

To improve solubility, solvent-accessible surface areas of side chains in the TrpRS Urzyme were compared with corresponding side chains in native full-sized TrpRS in three different conformations, the open unliganded state, the pre-transition state, and the closed, liganded product state, using Areaimol (21). A threshold increase of 50 \AA^2 in exposure of hydrophobic residues caused by CP1 and CTD removal was used to identify residues that would be subjected to mutation by Rosetta design. Twelve predicted mutations were introduced to the WT TrpRS Urzyme in addition to those previously selected for CP1 loop closure to construct the redesigned (DES) Urzyme. Genes for both TrpRS Urzyme amino acid sequences were synthesized by Genscript, who eliminated rare codons from the synthetic genes.

Subcloning—Restriction enzymes and other molecular biology reagents were obtained from New England Biolabs and were used according to the manufacturer's protocols. Forward and reverse primers for PCR amplification were designed to be complementary to the Urzyme sequence, containing extra nucleotides coding for N-terminal FLAG and C-terminal His₆ tags. Desired restriction sites were also introduced in the primer design step. Following digestion with the appropriate restriction enzymes, the PCR products were cloned to *Esche-*

TrpRS Urzyme Kinetics

richia coli expression plasmids pET SUMO (Novagen) or pMAL-c2x (New England Biolabs) for use in bacterial strain BL21(DE3)pLysS (Novagen). All constructs were confirmed by DNA sequencing.

Expression and Purification of MBP Fusion—MBP-TrpRS Urzyme fusion was expressed directly after transforming with plasmid DNA. After transformation, 100 μ l of competent cells were incubated in 1 ml of LB media at 37 °C for 1 h with shaking and then used to inoculate 100 ml of growth media (LB plus 2% glucose, 100 mg/ml ampicillin). The inoculum was shaken at 250 rpm overnight at 37 °C and then transferred to 1 liter of fresh LB with 100 mg/ml ampicillin. At $A_{600} = 0.4 - 0.6$, isopropyl β -D-thiogalactopyranoside was added to 0.3 mM final concentration to induce gene expression. Harvest began 3 h after induction.

Cells were collected by centrifugation at 4,500 rpm for 20 min at 4 °C and then resuspended in lysis buffer (50 mM Tris, 10% sucrose, pH 7.5). After overnight storage at -80 °C, the cells were thawed on ice and lysed by sonication in pulse mode (three times for 15 s). The lysate was centrifuged at 14,000 rpm for 30 min at 4 °C to separate the supernatant fraction from the inclusion bodies. The supernatant from this step was diluted 1:5 with the amylose resin High Flow (New England Biolabs) column buffer (20 mM Tris-HCl, 0.2 M NaCl, 1 mM EDTA, 10 mM β -mercaptoethanol, pH 7.4) before loading onto a disposable column (Bio-Rad) containing 2 ml of amylose-conjugated resin suspension for each 10 ml of the diluted supernatant. The column was equilibrated with 3 column volumes of column buffer, and the resin was washed with 5 column volumes. The flow-through fraction was collected, and then the resin was washed with 8 column volumes of column buffer to minimize nonspecific binding. The bound fusion protein was eluted in 1.5-ml fractions with elution buffer (column buffer plus 10 mM maltose). The first five fractions were combined and dialyzed three times overnight against 10 volumes of 50% glycerol dialyzing buffer (20 mM Tris-HCl, 50 mM KCl, 0.1 mM PMSF, 10 mM β -mercaptoethanol, 50% glycerol, pH 7.5) at 4 °C. Samples were stored at -20 °C after dialysis.

Factor Xa Cleavage—The MBP fusion was digested with Factor Xa (New England Biolabs) overnight at 4 °C with rotation. The ratio of protease to fusion protein is 1:50 (v/v). Concentrated fusion protein was diluted to a final concentration of 10% glycerol from stock stored in 50% glycerol before digestion. The digestion mixture was used directly for tryptophan activation assay.

Active-site Titration—The concentrations of active sites, compared with the total concentrations of DES TrpRS fusion protein and Urzyme, were determined using the active-site titration procedure as described previously (22) using 10 mM [γ - 32 P]ATP in excess over the total estimated enzyme concentration (1.7 mM). Reactions were quenched with SDS at time intervals between 3 and 45 min and aliquots examined using thin layer chromatography on polyethyleneimine cellulose. After development, plates were scanned using a Typhoon phosphorimager. Amounts of labeled ATP and PP_i were estimated from densitometry and fitted to Equation 1 using JMP (23).

$$[\text{ATP}] = A \cdot \{\exp(-k_{\text{chem}} \cdot t) - k_{\text{cat}} \cdot t + C\} \quad (\text{Eq. 1})$$

The active fraction is obtained from the fitted parameters (24) as shown in Equation 2.

$$n = 1.07 \cdot (1/(1 + C)) \cdot [\text{ATP}]_{\text{total}} / [\text{enzyme}]_{\text{total}} \quad (\text{Eq. 2})$$

ATP- and Tryptophan-dependent 32 P $_i$ Exchange Assays for Tryptophan Activation—Samples were assayed using the traditional 32 P-pyrophosphate exchange as described previously (25, 26) 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 μ l of pyridine at 37 °C (19). ATP and tryptophan were first depleted from the assay buffer to determine ATP- and tryptophan-dependent Michaelis-Menten constants. Stocks of ATP and tryptophan were added to the depleted buffer at indicated concentrations. Activity measurements and substrate concentrations were fitted to the Michaelis-Menten equation using JMP (23). The k_{cat} value was based on the estimated concentration of TrpRS Urzyme released after 80% efficient Factor Xa digestion, using the ImageJ image-processing program (National Institutes of Health) to obtain the relative densities of proteins on a Coomassie-stained SDS-polyacrylamide gel and corrected for the fraction of active sites.

Molecular Dynamics Simulations—The structure output by Rosetta for the DES TrpRS Urzyme was used to construct the initial model for MD simulations using the Sigma program. The final system contained 2157 protein atoms, including hydrogen atoms and 4684 water molecules, described by the CHARMM force field and TIP3P water model, respectively, as described previously (27).

RESULTS

As we were interested in more radical protein surgery necessary to remove the CP1 peptide and the anticodon-binding domain, our approach should be distinguished from ancestral gene resurrection. Information about possible amino acid sequences of long extinct prototypic enzymes necessary to test the Rodin and Ohno hypothesis has vanished. The challenge therefore lies in identifying compensatory amino acid mutations in the remaining polypeptide to create a model for the ancestral enzyme. Those sequences cannot be assumed to resemble contemporary proteins because large masses must be deleted, and in the case of the TrpRS Urzyme, removing these masses cleaves several of the chief hydrophobic cores from which the contemporary protein derives its stability. Moreover, exposure of the cleaved nonpolar cores leads inevitably to aggregation and low solubility.

TrpRS is, nonetheless, a nearly ideal system with which to attempt this radical protein surgery. It is the smallest aaRS, and the gap between the two free ends resulting from CP1 removal is only 5 Å long, which can be joined directly by relaxing the adjacent backbone. We used the biophysical chemistry implemented in the protein design program Rosetta (20) to approximate sequence changes involved in the putative evolutionary process, which have faded from the amino acid sequence phylogenies.

Computational Design Ensures TrpRS Urzyme Molecular Continuity, Stability, and Solubility—Two genes encoding the highly conserved residues in two disjoint fragments of the

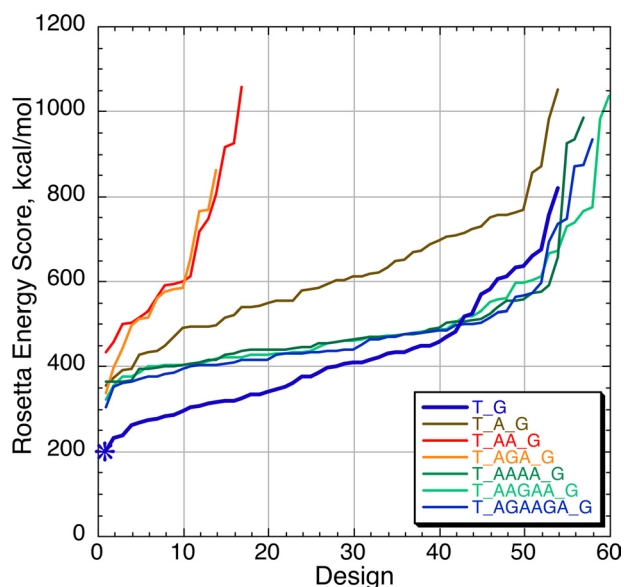


FIGURE 1. Rosetta energy scores are decisively lower for the 130-residue Urzyme with a zero length linker (direct linkage between Thr-46 and Gly-121) and five relaxed side chains adjacent to the site of CP1 removal. Multiple designs for each of six other possible linker lengths are shown in colors ranging from red (the worst). The designed sequence indicated by the blue asterisk is the subject of this work.

TrpRS Rossmann fold were constructed (19) based on multiple structure alignments of known class I aaRSs. In both these 130-amino acid Urzymes, the connecting peptide (CP1) was removed; the N- and C-terminal crossover connections of the Rossmann fold were ligated, and the C-terminal domain was truncated at residue 204.

These deletions exposed the molecule to two anticipated problems. The first of these was structural disruption at the site of CP1 removal. To reconnect the two disjointed active-site fragments, we used the Rosetta LOOP algorithm to substitute short data base-derived linkers of varying length, from 0 to 6 residues, containing the short side chain amino acids alanine and glycine to minimize interactions between the inserted side chain and the main backbone. These were scanned to identify optimal backbone continuity between the N- and C-terminal fragments of the Urzyme (Fig. 1). Furthermore, four residues adjacent to sites Thr-46 and Gly-121, where CP1 was removed, were allowed to mutate (with backbone relaxation), facilitating the joining of two free ends. This construct also had five mutations, illustrated by the line of blue spheres at the site of removal (Fig. 2B).

It is notable that of all seven linkers tested, ranging from 0 to 6 amino acids, the direct linkage between Thr-46 and Gly-120 was consistently predicted by Rosetta to be most stable (Fig. 1). This result is consistent with the presumed existence of an ancestral class I aaRS lacking the connecting peptide 1, and suggests that CP1 was a later insertion.

More severe problems were instability and/or insolubility due to the loss of two-thirds of the stabilizing mass afforded in the full-length TrpRS. Disruption of nonpolar core packing is exacerbated because in class I aaRS the region forming the chief hydrophobic core in many canonical Rossmannoid proteins has been evacuated to accommodate the amino acid-binding site (Fig. 2A). Thus, a significant portion of the TrpRS hydrophobic

core resides in the interface between the Urzyme and CP1 and becomes exposed by removing CP1.

An exception is a 7-residue core at the C terminus of the N-terminal α -helix of the Rossmann fold. We have noted that this core motif likely participates in conformational switching and transition state stabilization in native TrpRS (28). Notably, this motif remains intact in the TrpRS Urzyme.

Elsewhere, however, deletions made in constructing the Urzyme uncovered a large number of buried nonpolar side chains. Because of this problem, we were unable to refold the WT TrpRS Urzyme without generating extensive aggregation, as evidenced by heavy precipitation as the 6 M urea was dialyzed away (19).

To address the instability and insolubility resulting from exposing core side chains, the profile of the accessible surface area for each residue in the computationally simulated Urzyme was compared with the native enzyme using Areaimol (21). From 60 residues with increased surface area after removal of CP1 and the CTD, the 19 residues with the greatest percentage increase in surface area were selected, and Rosetta was run to identify the mutations (Table 1).

Of those 19 residues, 12 mutations predicted by the design program to improve solubility and stability were included in a second construct. Except for Tyr-91, these mutations all were made to nonconserved residues. The two TrpRS Urzyme constructs in Fig. 2, designated wild type (WT) and redesigned (DES), were synthesized after the two phases of linker design and stability-solubility optimization (19).

Rosetta design has a tendency to place patches of hydrophobic residues together (29). Thus, its ability to improve Urzyme solubility was somewhat less than its ability to recover stability and proper folding. Nevertheless, of the mutations predicted to stabilize and solubilize the DES TrpRS Urzyme, five are from hydrophobic to hydrophilic (Fig. 2B) and would directly improve the interaction with solvent.

The redesigned TrpRS Urzyme also expressed predominantly in inclusion bodies. However, in contrast to the WT Urzyme, we were readily able to refold it from 6 M urea by successive dialysis to remove (19) and to demonstrate its catalytic activity. Thus, Rosetta afforded a decisive advantage in stability and, presumably, correct folding (19).

SUMO Expression System Improved Yield but Not Solubility, although the Majority of MBP Fusion Protein Was in the Soluble Fraction—Both WT and DES TrpRS Urzymes were insoluble when expressed as pET42 constructs. Subcloning the two genes into Invitrogen pET small ubiquitin-related modifier (SUMO) vector improved the expression level, but the SUMO fusions stayed insoluble in the pellet (Fig. 3A, P). The size of SUMO-fused redesigned Urzyme estimated by SDS-PAGE was at the expected position of \sim 25 kDa. TrpRS Urzyme accounted for 56% the mass of the fusion, which suggested that the 11-kDa SUMO carrier may be too small to overcome the insolubility of the Urzyme. No effort was made to measure enzymatic activity from the SUMO fusions.

Maltose-binding protein (MBP) is approximately three times larger than the subcloned open reading frame of TrpRS Urzyme. A linearized pMAL-c2x map used for subcloning is shown in Fig. 3B. The construct was doubly tagged with the

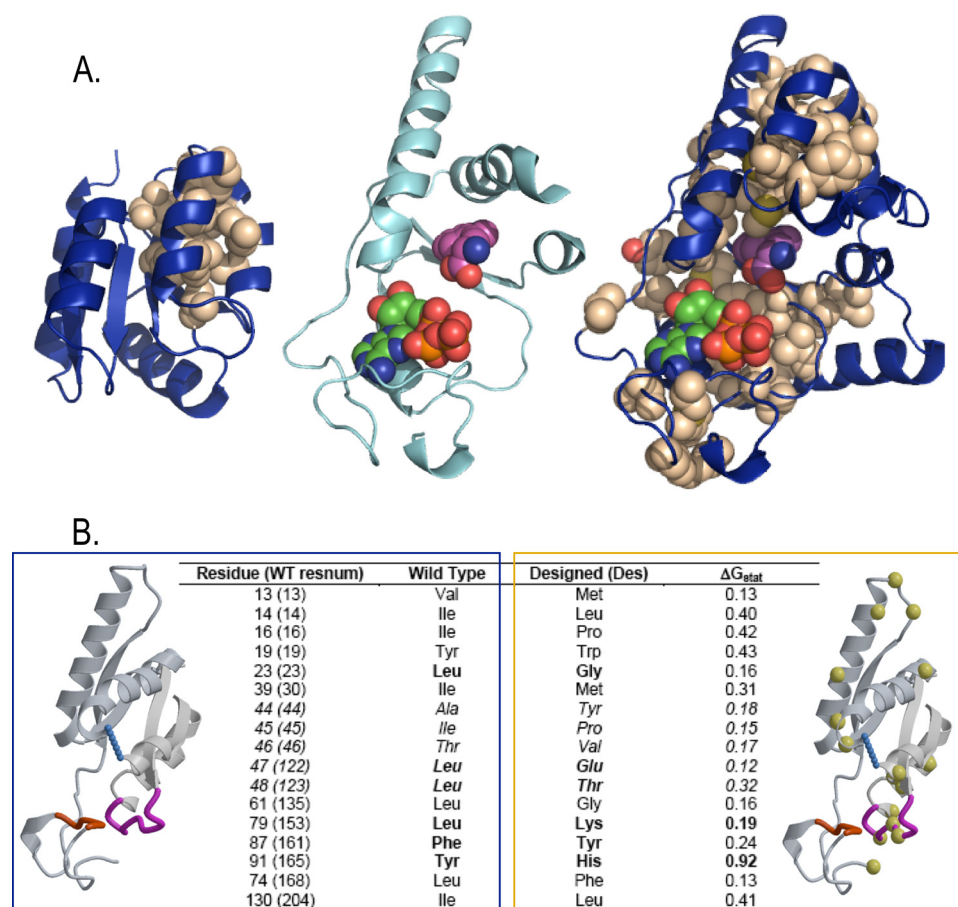


FIGURE 2. *A*, comparison of nonpolar cores in TrpRS Urzyme (*center*) and catalytic domain (*right*) with the canonical Rossmannoid homolog, FixJ (*left*). All structures were superimposed using the 46 residues of the N-terminal α - β - α crossover connection (42). Note the location of the two substrates, ATP (*green*) and tryptophan (*magenta*). Tryptophan occupies a region which, in FixJ, contains a substantial hydrophobic core. Migration of the core away from the interface between the central β -sheet and the helix to accommodate the amino acid substrate heightens difficulties posed by stripping away the CP1 peptide and anticodon-binding domain, making protein design more critical to class I Urzyme construction. *B*, schematic of mutations introduced to the TrpRS Urzymes, including five residues that were changed from hydrophobic to hydrophilic in the WT and DES Urzymes, respectively (*boldface*). Free energy differences were also calculated showing minimal effect of the mutations. A ΔG_{stat} of less than 0.45 is a measure of relative sequence conservation as defined by Lockless and Ranganathan (60). The redesigned Urzyme harboring 12 mutations (Table 1) is predicted to improve stability and solubility (adapted from Pham *et al.* (19)).

TABLE 1

Solvent-accessible surface area between WT Urzyme and full-length TrpRS

Three different configurations are compared as follows: PROD, Trp-5' AMP complex; Pre-TS, pre-transition state complex to ATP and tryptophanamide; open, unliganded.

Amino acid	Residue	$\Delta S(\text{Prod})$	$\Delta S(\text{Pre-TS})$	$\Delta S(\text{open})$
Val	13	9	99	19
Ile	14	85	71	102
Ile	16	57	62	58
Tyr	19	98	74	83
Leu	23	57	56	63
Tyr	36	90	86	65
Phe	37	84	88	95
Ile	39	124	127	124
Leu	122	132	120	89
Leu	123	115	119	117
Leu	135	130	119	112
Tyr	136	87	81	91
Leu	153	64	58	56
Leu	157	52	55	51
Phe	161	72	75	76
Tyr	165	89	80	74
Leu	168	75	76	69
Leu	186	105	91	109
Ile	204	114	119	134

N-terminal FLAG and C-terminal His₆. Immunoblots of soluble (supernatant) and insoluble (pellet) fractions with anti-FLAG antibody showed that the majority (80%) of the fusion was in the supernatant (data not shown). Affinity chromatography on amylose resins gave pure, soluble fusion protein (Fig. 3*B*). Because of the low catalytic activity of the Urzyme, data such as that in Fig. 3*B* do not adequately describe purity with respect to very minor contaminating amounts of native *E. coli* TrpRS, which must be addressed by other controls. These are summarized below.

The MBP fusion protein appeared to be inactive in the initial tryptophan activation assay (data not shown). Factor Xa cleavage of the fusion protein released about 80% of the Urzyme from the MBP fusion. Several assays on inactive cleaved preparations showed no activity, ruling out activity arising from Factor Xa. The digestion mixture was used directly for ³²P_i exchange assays, showing that digestion also released a cryptic tryptophan activating activity from the fusion protein.

A major concern has been the degree to which the Urzyme constructs fold properly. We used this activity to address that question here as follows: first by carrying out active-site titration (24) of both fusion protein and Urzyme, second by Michaelis-Menten steady-state assays and determination of relative specificity for tryptophan *versus* tyrosine, and finally by carrying out long molecular dynamics simulations. Many factors likely impact the activity and properties of the TrpRS Urzyme, and we have managed only to control some of these. Thus, all experimental measurements are subject to high variance. Thus, it is difficult, for example, to compare the experiments that follow on a quantitative basis because they have been done with different Urzyme preparations. Nevertheless, we document here three conclusions that lie well beyond experimental error. (i) Active-site titration experiments show that previous estimates for the Urzyme activity are lower limits and that the actual activity is higher, because of the fact that only ~35% of the molecules are active. (ii) Michaelis-Menten experiments show that both k_{cat} and the affinity for ATP are close to corresponding values for the full-length enzyme. (iii) Weaker amino acid affinity in the Urzyme leads to significantly reduced amino acid specificity.

Nearly 90% of the surface area buried in the full-length TrpRS dimer arises from the deleted CP1 peptide. On this basis, we expect that the Urzyme is active as a monomer. Attempts to confirm this conclusion definitively have been inconclusive, because gel filtration patterns of both the MBP fusion protein and the cleaved products under native solution conditions are broad and of elevated molecular weight, implying nonspecific aggregation.

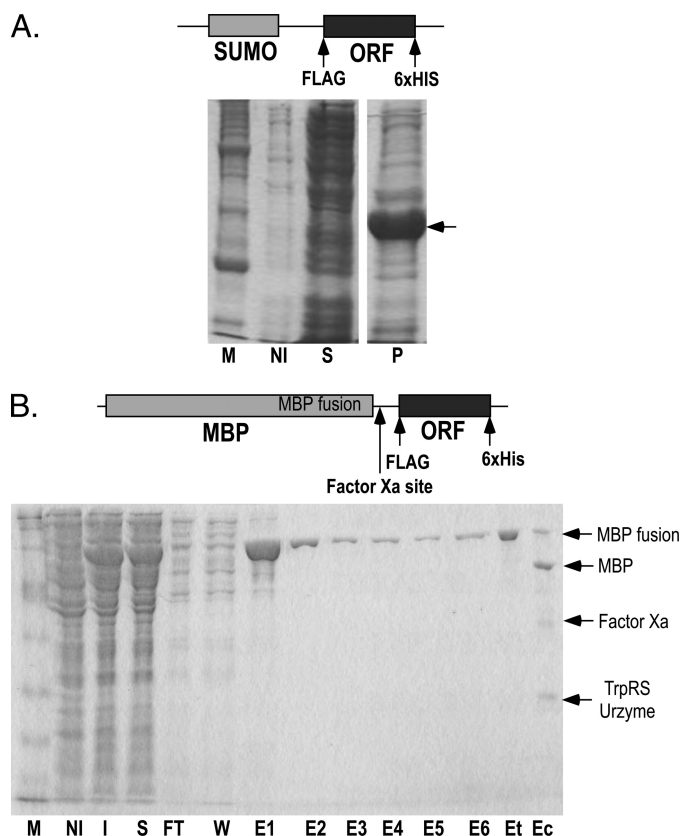


FIGURE 3. *A*, SUMO expression system increases total Urzyme production but not solubility. *M*, molecular weight markers; *NI*, noninduced control; *S*, supernatant; *P*, pellet. *B*, expression and purification of maltose-binding protein-fused redesigned TrpRS Urzyme. *I*, induced expression; *S*, soluble fraction; *FT*, flow-through fraction from amylose column; *W*, 1st wash after flow-through fraction; *E1–6*, eluted fractions; *Et*, concentrated fusion; *Ec*, Factor Xa cleavage products.

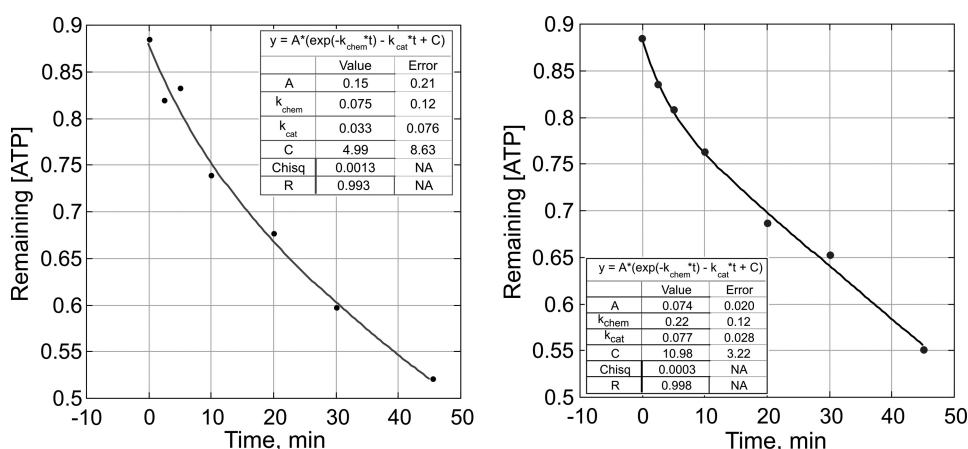


FIGURE 4. Active-site titration of DES TrpRS fusion protein (left) and Urzyme (right) expressed at 18 °C. The [ATP] and [protein] were 10 and 1.7 mM, respectively. Both curves exhibit a pre-steady-state burst, from which the fraction of active molecules can be estimated (24).

Active-site Titration Clarifies the Relative Activities and Folding of the DES Fusion Protein and Urzyme—Active-site titration experiments provide several key clarifications. First, they confirm that even after such drastic surgery on the native TrpRS structure, product release remains rate-limiting for the Urzyme. This was certainly not a foregone conclusion, as removal of an entire domain and the CP1 subdomain could have abolished the ability of the Urzyme to retain the adenylate intermediate and hence to exhibit a pre-steady-state burst. The titration experiments show convincingly that substantial bursts occur (Fig. 4).

Second, titrations provide estimates of the degree of proper folding. The algorithm described by Fersht *et al.* (24) affords an estimate that about 81% of the fusion proteins and 35% of the cleaved Urzyme preparations have competent active sites. These estimates narrow the range of values for k_{cat} reported below. Absent the active-site titrations, values we obtained would necessarily only be lower limits.

Finally, titrations confirm that the fusion protein is inhibited by the MBP tag. Thus, although the fusion protein has a higher fraction of active sites, its k_{chem} value is about a third that of the cleaved Urzyme, in keeping with our recurrent inability to demonstrate steady-state activity in fusion protein preparations.

TrpRS Urzyme Tryptophan Activation Activity Depends on Both ATP and Tryptophan—Michaelis-Menten experiments substantiate the use of the TrpRS Urzyme to represent ancestral aaRS activity (Fig. 5). There was no activity in the absence of either tryptophan or ATP. The specific activity previously estimated from the specific activity of refolded DES TrpRS Urzyme expressed in pET42a plasmid was $\sim 10^5$ times weaker than that of the full-length enzyme (19). As a result of the low activity, steady-state kinetics experiments reported here for the fusion protein and cleaved products were noisy, and titrations with both substrates were replicated multiple times.

Fig. 5 and Table 2 show that the DES TrpRS Urzyme kinetics differ significantly from those of full-length TrpRS. Moreover, the details themselves are interesting. The k_{cat} ($\sim 4.9 \text{ s}^{-1}$) and K_m values of Urzyme for ATP (0.4 mM) are quite similar to those of the full-length enzyme. The K_m value for tryptophan observed for Urzyme induced at 37 °C, 4.8 mM, is ~ 2000 times weaker. Indeed, when k_{cat} is corrected for the active fraction, the tryptophan K_m value is by far the most significantly different between the full-length TrpRS and its Urzyme.

Adventitious Contamination Cannot Account for the Observed Activity—Consistent with our earlier report (19), steady-state kinetics (Table 2) confirm that the catalytic activity of the TrpRS Urzyme is a 4- or 5-order of magnitude weaker catalyst than the full-length enzyme. Under these circumstances, one must do everything possible to ensure that the observed activity arises from the Urzyme and not

TrpRS Urzyme Kinetics

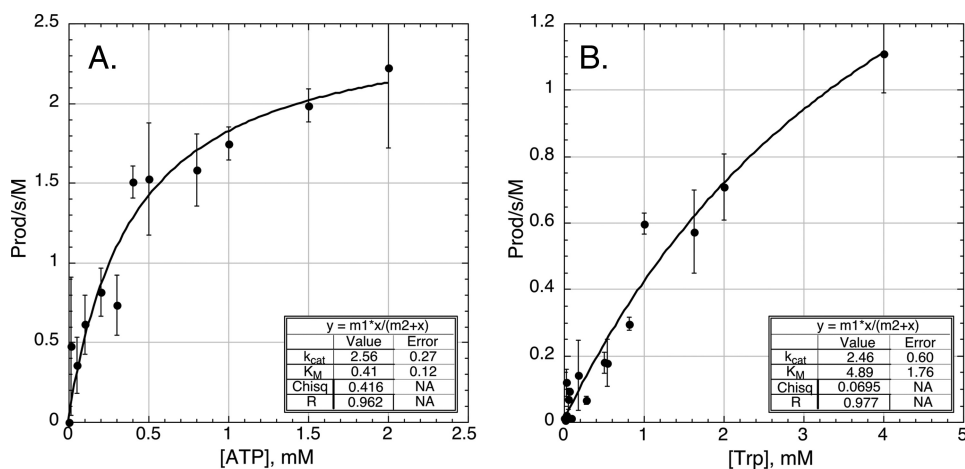


FIGURE 5. Steady-state kinetic analysis of $^{32}\text{PP}_i$ exchange activity of TrpRS Urzyme expressed at 37 °C. A, ATP dependence is similar to that of full-length TrpRS. B, tryptophan dependence does not saturate at accessible tryptophan concentrations. Error bars show standard deviations of measurements made at the same substrate concentrations. Error estimates for k_{cat} and K_M are indicated in the insets. Prod, product.

TABLE 2

Kinetic parameters of full-length TrpRS and its Urzyme

Enzyme parameters	TrpRS enzyme	TrpRS Urzyme	Urzyme:full length
$K_{m,ATP}$	0.0004 ± 0.02 M	0.00041 ± 0.12 M	~ 1 M
$k_{cat,ATP}$	27.2 s $^{-1}$	4.88 ± 9.5 s $^{-1}$	0.18 s $^{-1}$
k_{cat}/K_M	68,000 Prod	11,902 s	0.18 M
$K_{m,Trp}$	$2.00 \text{ E}-06$ M	$2.41 \text{ E}-3 \pm 2.1$ E-3 M	1205 M
$k_{cat,Trp}$	26.5 s $^{-1}$	4.69 ± 10 s $^{-1}$	0.18 s $^{-1}$
k_{cat}/K_M	13.2 E6 Prod	1.94 E3 s	$1.5 \text{ E}-4$ M

from adventitious contamination. Co-purification of as little as 1 in 10^4 – 10^5 of chromosomal *E. coli* TrpRS could account for the observed activity. The following experiments strongly support the conclusion that the activity reported here and earlier (19) represents authentic catalysis by the Urzyme. (i) A pMAL-c2x empty vector control extract purified by the same procedure has negligible activity nor was there activity in several unsuccessful efforts to purify the fusion protein, consistent with the absence of contaminating wild type activity. (ii) Purifying and renaturing the Urzyme from the insoluble inclusion body fraction tends to eliminate contaminating soluble host cell activity (19). (iii) Cryptic activity is released from the fusion protein after Factor Xa cleavage (Fig. 4). (iv) Active-site titration reveals that activity arises from a significant fraction (35%) of the potential active sites in the cleaved Urzyme sample. This would clearly not be the case if a tiny fraction of contaminating full-length enzyme were responsible for the observed steady-state activity. (v) The active site D146A mutation has opposite effects on full-length TrpRS and the Urzyme. The mutation reduces the activity of the full-length enzyme 200-fold, but it increases that of the Urzyme 25-fold (19). (vi) The $K_{m,Trp}$ value for urzyme (Fig. 5 and Table 2) is ~ 2000 -fold higher than that of the full-length enzyme. If contaminating *E. coli* TrpRS were responsible for the activity, it would saturate at ~ 2 μM , not at much higher tryptophan concentrations.

TrpRS Urzyme Has a Structure Similar to That Observed in the Full-length Enzyme—As yet, we have limited structural data for the TrpRS Urzyme. The redesigned Urzyme still has low solubility when cleaved from the MBP fusion. However, the tertiary structure derived from Rosetta design is stable in 70-ns

molecular dynamics calculations (Fig. 6, A and B), suggesting that the observed catalytic activity derives from a structure similar to that expected from the structure of the full-length enzyme from which it is derived. Moreover, as described more fully below, MD simulations help rationalize the significant differences in Trp affinity between the Urzyme and full-length TrpRS.

High K_M Value of Urzyme for Tryptophan Implies Reduced Amino Acid Specificity—Although the residues lining the ATP and tryptophan binding pockets are nearly all preserved in the Urzyme, the tryptophan affinity is $\sim 10^3$ -fold weaker. Weak amino acid binding strongly

suggests that the Urzyme has significantly reduced amino acid specificity, relative to the contemporary enzyme. To test this notion, we examined the specificity ratio, $(V_{max}/K_M)_{Trp}/(V_{max}/K_M)_{Tyr}$ (Table 3). The Urzyme is ~ 500 -fold less specific than full-length TrpRS for tryptophan relative to the related amino acid tyrosine. This result suggests a rather unexpected view of the ancestral TrpRS Urzyme, in which ATP bound tightly but tryptophan had only modest affinity.

A possible explanation is that longer range stabilizing interactions eliminated by removing the connecting peptide 1 and perhaps also the C-terminal anticodon binding domain significantly decrease an unfavorable entropy change necessary for amino acid binding to the Urzyme. Consistent with this hypothesis, long MD trajectories of the Urzyme complexed with Mg^{2+} ·ATP with and without tryptophan (Fig. 6B) reveal that the native-like orientation of the D helix, which we have referred to as the “specificity-determining helix” because it provides several residues crucial to specific amino acid recognition, depends strongly on bound tryptophan. In the absence of tryptophan, this helix deviates quite far from its position in the native enzyme, leading to a divergent trajectory. The presence of bound tryptophan preserves its native-like orientation.

Fig. 6C highlights the fact that CP1 surrounds and constrains the orientation of the specificity-determining helix. It appears to serve as an exo-skeleton that stabilizes the configuration of the tryptophan binding pocket. The MD trajectories therefore provide a persuasive rationale for the observed weakening of amino acid binding by the Urzyme and the consequent loss in fidelity.

DISCUSSION

The TrpRS Urzyme represents achievements in protein biophysics and biochemistry, and its measurable catalytic activity establishes a basis for experimental studies of the very early evolution of biological catalysis and molecular biology. We begin by discussing how *in silico* protein design used to construct the TrpRS Urzyme (19) extends previous efforts to dissect catalytically active domains from native, full-length pro-

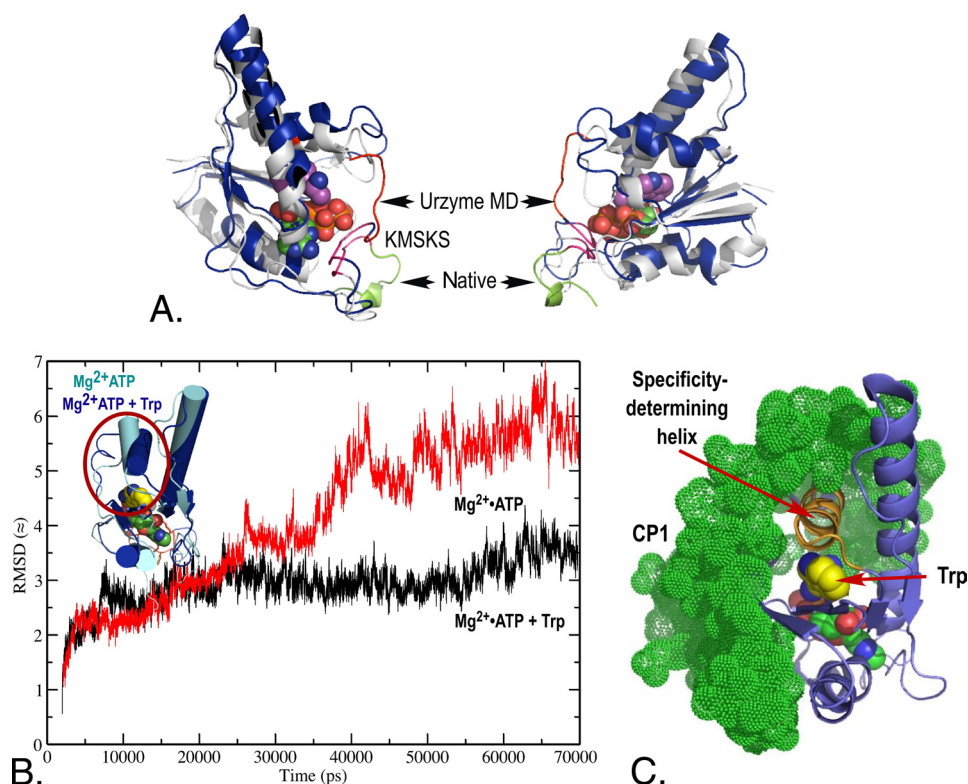


FIGURE 6. *A*, superposition of an MD snapshot of the TrpRS Urzyme after 1200 ps (blue) on the corresponding pre-transition state structure derived from Protein Data Bank code 1MAU and complexed to both substrates. The divergent segments are limited to the extreme C terminus following the catalytic KMSKS signature, which aligns closely. *B*, 70-ns MD simulation trajectories of the Mg^{2+} -ATP complex with and without tryptophan. The presence of tryptophan stabilizes the Urzyme in a conformation close to that of the pre-transition state structure (black), although ATP alone favors a conformation closer to the open state of the full-length TrpRS (Protein Data Bank code 1MAW). The inset highlights differences in the amino acid binding pocket between the two trajectories. Re-orientation of the specificity-determining helix in the absence of tryptophan is noteworthy. *C*, representation of CP1 (green spheres) wrapping around the core TrpRS Urzyme. Most interactions occur with the amino acid binding pocket containing tryptophan (yellow). *RMSD*, root mean square deviation.

TABLE 3
Specificity ratios of full-length TrpRS and its Urzyme

	V_{max}/K_m	Specificity ratio
WT (Trp)	93,478	4328
WT (Tyr)	21.6	
Urzyme (Trp)	547.1	10.0
Urzyme (Tyr)	54.9	

teins, and then consider mechanistic and evolutionary implications of the Urzyme's kinetic parameters.

Urzyme Construction Methods Substantially Extend the Experimental Study of Mechanistic Enzymology and Molecular Evolution—Rosetta design was used here as a source for missing phylogenetic data in TrpRS Urzyme construction. Ancestral gene resurrection utilizes multiple sequence alignments to deduce probable sequences for ancestral nodes. Because we were interested in an analogous reconstruction from a much more ancient period, phylogenetic data are insufficient to infer sequences that might have stabilized ancestral fragments like the TrpRS Urzyme. Nonetheless, natural selection for stability (30) was a likely determinant of biological sequence alignments. To that extent, Rosetta energy functions afford a sensible surrogate for information missing from multiple sequence alignments, extending both the time frame and the modularity

accessible to the experiment. In this way, protein design likely will enhance what generally can be done by protein surgery.

Previous efforts to generate active fragments from contemporary proteins have led to very diverse constructs that retain varying levels of catalytic activity. We referred to the 130-residue TrpRS Urzyme as a “minimal catalytic domain” (19). At the other extreme, a 527-residue fragment of acetylglucosaminyltransferase V (~70% of the full-length enzyme) was also referred to as a minimal catalytic domain (31). We have coined the term “Urzyme” to distinguish active constructs that retain only sufficient mass to position the active-site residues for catalysis.

The domain structures of aminoacyl-tRNA synthetases motivated previous experimental truncations to isolate active catalytic fragments (32–35). These efforts focused mainly on cleaving between the catalytic and anticodon-binding domains, a widely documented approach applied routinely only to proteins whose domain structures can be readily identified (36–38). For example, an N-terminal 320-residue HisRS fragment (~60% of the

native monomer) that retains an intact catalytic domain is 10^3 -fold less active without its deleted anticodon-binding domain (32). Such truncations are far less aggressive than those made to cognate tRNAs, which ultimately resulted in identifying mini-helices ~30% the size of intact tRNAs (39, 40) that could be charged by several aaRS. At 40% of its native molecular weight, the TrpRS Urzyme represents a comparable truncation for an aaRS. Recapitulation experiments restore modular information, such as CP1, which was removed to construct the Urzyme.

The Urzyme construction also resembles the joining of domains from two related enzymes with different activities to make an active fusion (41). The N-terminal fragment contains the TIGN signature, together with a highly conserved core packing motif (42) that we have implicated in conformational switching (43) and activation of the catalytic Mg^{2+} ion (28) in full-length TrpRS. The C-terminal fragment contains residues 125–137, which we previously termed the specificity-determining helix, as well as the GXDQ and KMSKS catalytic signatures. The size of the resulting TrpRS Urzyme thus seems minimally sufficient to position the active site residues.

Schwob and Söll (33) provided the most direct precedent for this work. They selected *in vivo* for active truncations of the class I aaRS GlnRS, developed by separate and nested deletions of either the CP1 insertion or the anticodon-binding domain,

TrpRS Urzyme Kinetics

but not both. Their work demonstrated that various fusions of the first and second halves of the Rossmann dinucleotide binding fold could preserve *in vivo* activity. Here again, our work breaks new ground by deleting an entire CTD and fusing two disjoint halves of the ATP-binding Rossmann fold after removing the 74-residue intervening CP1 peptide. Notably, in light of our own similar difficulties, the GlnRS constructs could not be assayed *in vitro* because they aggregated.

We note three important differences between the TrpRS Urzyme and these GlnRS constructs. First, TrpRS Urzyme is 130 residues, compared with ~460 residues in the minimal GlnRS fragments. Second, the GlnRS CP1 deletions left intact B- and C-helices with the intervening β -strand, residues 71–115 (residues 49–96 in TrpRS), which are deleted entirely in the TrpRS Urzyme without losing *in vitro* activating activity! Finally, although both GlnRS constructs and the wild type Urzyme aggregated irreversibly, redesign allowed us to overcome the irreversible aggregation observed for both the GlnRS constructs and the wild type Urzyme. Factor Xa cleavage of the MBP fusion necessary to release the inhibited activity also led occasionally to aggregation of the cleaved Urzyme, which was exacerbated after longer incubation times, at higher temperature, and in the absence of glycerol. Aggregation remains a problem to be addressed by further studies.

Accumulation of expressed WT and DES Urzymes in inclusion bodies may contribute to the rapid loss of Urzyme-containing plasmids, even when selective pressure is maintained by using antibiotics. This apparent toxicity persists when the Urzyme is expressed in BL21(DE3) pLysS, an expression system that should minimize basal expression and hence toxicity. Thus, some function, perhaps misacylation, probably occurs *in vivo* and contributes to plasmid toxicity. Further work is necessary to answer such questions.

The TrpRS Urzyme has potential applications for customized protein synthesis, which is of considerable interest (45, 46). Suga and co-workers (47) have pioneered the use of ribozymal acylating catalysts to incorporate noncanonical amino acids into proteins for the purpose of protein engineering. Their ribozyme, however, cannot activate amino acids. High catalytic activity and low specificity make the TrpRS Urzyme a potentially useful complement to this “flexizyme.” Functional complementation between polypeptide and RNA catalysts in such a context could model an important bridge between putative prebiotic RNA- and ribonucleoprotein-world scenarios.

Urzyme Enzymology (Urzymology) Reveals Cryptic Intramolecular Interactions Necessary for Specificity and Catalytic Activity—The 74-residue fragment 46–120 we refer to here as CP1 previously was considered integral to the Rossmann fold. The relative independence of TrpRS Urzyme function from the three helices suggest altering the definition of the CP1 insertion from that originally proposed (1, 13) to also include this fragment. Indeed, CP1 is frequently used to refer only to the domain structures suspended from a long, anti-parallel β -arm that is conserved only in class Ia and Ib aaRS. However, the 5-Å distance between the first and second crossover connections of the Rossmann fold is distinctly conserved in all class I aaRS.

The significant impacts of removing CP1 and the CTD on amino acid activating activity provide compelling evidence for

functional contributions of these modules in the full-length enzyme. Our previous mutagenesis of the active-site lysine residues in full-length TrpRS (28) suggest that rotation of the CTD endows the catalytic Mg^{2+} with a conditional role in catalysis. The conditional use of metal ion catalysis ensures that catalysis of amino acid activation is coupled to domain movements necessary for tRNA aminoacylation. Although the switching residues we have identified in the N-terminal crossover connection are retained in the TrpRS Urzyme, it is likely that addition of the CTD allowed the evolutionary development of a more sophisticated allosteric mechanism.

Parameters in Table 3 suggest that these contributions also enhance substrate specificity, particularly the *relative* affinity for different amino acid substrates. Contemporary full-length TrpRS is very specific for its cognate amino acid substrate, tryptophan (48), whereas the Urzyme is not. As noted above, CP1 surrounds the tryptophan binding pocket and forms nonpolar cores on its three faces. A key role of the amino acid in the Urzyme-catalyzed reaction appears to be to re-orient the specificity-determining α -helix in the catalytically productive complex, suggesting that much of the tryptophan-binding energy is used to reconfigure the amino acid-binding site in the Urzyme, consistent with its weaker affinity. It is tempting, especially in light of the location of class Ia editing domains within CP1, to relate the surprisingly weak tryptophan binding to the absence of CP1 (Fig. 5C). The inserted CP1 appears to work as an exoskeleton to stabilize the proper configuration necessary for tryptophan recognition.

The CP1 helices form nonpolar cores with three regions surrounding the tryptophan-binding site on the surface of the Urzyme. Their substantial hydrophobic surfaces sustain tertiary structure throughout the class I superfamily. As a result, removing CP1 creates a problem that is exacerbated by the fact that the amino acid binding pockets in class I aaRS are created at the expense of what in other Rossmannoid proteins is a substantial hydrophobic core (Fig. 2A). Stability of their tertiary structures, especially those that define the amino acid binding pocket, is thus necessarily exported to “outer sphere” core regions between the Urzyme and the rest of the protein.

Other published work suggests CP1-induced specificity enhancement derives from interactions between CP1 and the amino acid binding pocket in TrpRS and more generally in class I aaRS. Praetorius-Ibba *et al.* (48) attempted unsuccessfully to switch TrpRS specificity to favor tyrosine. Ten systematic mutants within the tryptophan binding pocket left the relative affinities for the two amino acids essentially unchanged. Their lack of success suggests that longer range interactions are necessary to stabilize the structure of the tryptophan binding pocket for high affinity. Thus, the range of mutations examined previously (48) may have more dramatic effects on amino acid discrimination by the Urzyme.

Similar attempts to change GlnRS binding specificity from glutamine to glutamate have confirmed the importance of longer range interactions in configuring the amino acid binding pocket (49, 50). As was observed for TrpRS, mutating only amino acids that interact directly with the amino acid is insufficient to switch specificity. However, engineering of two external loops and one deletion increase the generation of misacy-

lated Glu-tRNA^{Gln} by GlnRS 16,000-fold. Both engineered loops correspond to elements that interact closely with the C-helix (GlnRS residues 102–116; TrpRS residues 83–97), which is absent in the TrpRS Urzyme but present in the engineered GlnRS. Our observations therefore reinforce the conclusion of Bullock *et al.* (49) that high amino acid specificity is a distributed phenomenon and suggest that full recognition can arise only in the presence of CP1.

It is likely, because of the significant role played by the tRNA anticodon on tRNA recognition, that the CTD has a major impact on interactions with cognate tRNA (51–53). We are not yet in position to evaluate effects on tRNA specificity because we have not yet tested the TrpRS Urzyme for tRNA acylating activity. It lacks one tRNA determinant in the third helix of CP1 but retains residues 142–150 of the α E helix, which is the other element thought to include a tRNA acceptor-stem-binding site. Thus, as catalysis of second-order reactions generally involves reducing the entropy of activation, the Urzyme could possibly accelerate the second, acyl-transfer, reaction. It is notable in this respect that full-length TrpRS will actually acylate ATP, which behaves as the ultimate “mini-helix.” The assay described by Wolfson *et al.* (54) for tRNA acylation appears to have the requisite sensitivity to detect catalyzed acyl transfer at levels comparable with that observed for tryptophan activation.

Opposite Effects of the D146A Mutation in Full-length TrpRS and the Urzyme Imply Mechanistic Differences—ATP-dependent Michaelis-Menten parameters for the Urzyme appear to be quite similar to those of the intact enzyme. Thus, they are essentially independent of CP1 and the CTD. The surprising activation of the Urzyme by the D146A mutation (19) nevertheless implies that the Urzyme mechanism differs from that of full-length TrpRS. The highly conserved (42) core motif in the N-terminal 46 residues of the Urzyme (*i.e.* the D1 switch motif (43)), which also includes the ATP-binding site, was independently identified as an important allosteric activator of the catalytic Mg²⁺ ion in full-length TrpRS (28). Although this core is intact in the Urzyme, its function may differ substantially, because of the fact that the CTD, whose relative motion appears to be critical to the activation of the metal ion in full-length TrpRS, is entirely missing in the Urzyme. The two effects are possibly related. Absence of the CTD in the Urzyme could allow Asp-146 (Asp-68) to impede catalysis by stabilizing a nonproductive configuration of the metal ion (55). We suggest that the catalytic importance of the Asp-146 residue in intact TrpRS may be linked to the allosteric mechanism, which cannot function in the Urzyme.

These effects represent novel cryptic specificity and/or catalytic enhancements that have been revealed by study of the TrpRS Urzyme. The various putative enhancements provided by CP1 and the CTD to substrate recognition and catalysis can now be investigated experimentally by straightforward genetic construction, because the catalytic activity of the Urzyme itself can be measured. Similar experiments, in which the authors restored the truncated C-terminal domain *trans* and measured enhanced acylation activity, have been reported for GluRS (34). Using Rosetta to redesign the interfaces of free CP1 and CTD to the DES Urzyme should facilitate

testing their activity *in trans* to compare covalent *versus* non-covalent complementation.

Urzyme Kinetic Parameters Are Consistent with Those Expected for the Earliest Synthetases and Confirm a Key Prediction of the Rodin-Ohno Hypothesis—The TrpRS Urzyme accelerates tryptophan activation by $\sim 10^9$ – 10^{10} -fold, and this activity resides in the most highly conserved structural module of the class I aaRS superfamily. Such rate enhancement represents a substantial selective advantage over, for example, putative ribozymal catalysts (56). Application of Ockham's razor thus implies that ancestral class I aaRSs resembled the TrpRS Urzyme more closely than they did any radically different structure and that CP1 and the CTD were later additions that enhanced specificity, respectively, for cognate amino acids and tRNAs.

It was certainly essential that ATP be bound productively to primordial aaRSs. The native-like ATP binding affinity of the TrpRS Urzyme is thus consistent with its being a candidate for the ancestral synthetase. Moreover, there is considerable precedent for the relatively high ATP affinity for the Urzyme. Hol *et al.* (57) noted that phosphate-binding sites tend to occur at the N terminus of α -helices, because of the electrostatic potential of the four free amide nitrogen groups. Roughly 40% of the Urzyme mass consists of a core motif found to be remarkably intact in ~ 110 families of Rossmannoid proteins (42), and which contains, intact, the initial ATP-binding site of the native enzyme (58). Titration of ATP with the homologous 51-residue Walker A segment from the F₁-ATPase revealed that that peptide binds ATP with $K_d = 10 \mu\text{M}$ (59). We expressed and purified the corresponding TrpRS fragment, and it also binds ATP more tightly than does full-length TrpRS.³

By confirming a key prediction that catalytic activity is largely independent of CP1 and the CTD, the TrpRS Urzyme activity affords the first evidential support for the Rodin-Ohno hypothesis. Differences in structure and catalytic mechanism (*e.g.* tRNA acceptor stem-binding mode and hydroxyl group specificity) strongly support the consensus that class I and II aaRS had two distinct ancestors (2). However, translation of the reverse complements of nucleic acid sequences coding for the conserved motifs of one class shows distinctly nonrandom homology to the signature sequences of the other class (18), suggesting that their ancestral gene sequences were complementary.

Carter and Duax (61) produced circumstantial evidence for the sense/antisense coding hypothesis by noting that class I and II aaRS structures might resemble those of *Achlya klebsiana* glutamate dehydrogenase and HSP70 chaperonin (a stress protein), respectively. The latter two proteins are reported to be translation products from complementary strands of the same contemporary gene (62–64). This interesting observation generated controversial feedback as Williams *et al.* (65) recently challenged the conclusion that the *A. klebsiana* gene actually does code for the dehydrogenase. Further analysis of that gene, however (66), reinforces our original claim.

At first glance, the idea that the two synthetase classes originated from two complementary strands of one gene is very

³ L. Li, V. Weinreb, and C. W. Carter, Jr., manuscript in preparation.

unusual, because favorable mutations on one strand would likely be deleterious (e.g. lethal or nonsense) to the product encoded by the opposite strand. However, especially in a pre-cellular environment, coding both classes of aaRS classes on one gene would make efficient use of genetic material. More importantly, simultaneous expression of two synthetases with complementary specificities (19, 66) would favor relatively nonspecific translation of molten globular gene products from RNAs in their immediate environment. Only later, as the genetic code became firmly established and modern globular proteins assumed more specific structures, would it have been necessary to evolve higher amino acid specificity.

Multiple events of duplication, fusion, and insertion probably assembled the additional domains, enhancing substrate affinity, fidelity, and rate enhancements of the primitive enzymes. The remarkably idiosyncratic structure, mechanism, and accessory functions (4) of the 20 contemporary aaRSs resulted from this postulated divergent evolution. The measurable catalytic activity of the TrpRS Urzyme provides a key metric for experimental recapitulation of the evolutionary motif shuffling that likely served as a key mechanism to accelerate the evolutionary adaptation and functional improvement of primordial protein catalysts (67).

Acknowledgments—pMAL-c2x vector was a kind gift from A. Sancar. We thank H. Fried, N. Ozturk, L. Li, and O. Erdogan, for their assistance with the set up of procedures fundamental for the completion of this project.

REFERENCES

- Starzyk, R. M., Burbaum, J. J., and Schimmel, P. (1989) *Biochemistry* **28**, 8479–8484
- Carter, C. W., Jr. (1993) *Annu. Rev. Biochem.* **62**, 715–748
- Kirby, A. J., and Younas, M. (1970) *J. Chem. Soc. B* **418**, 1165–1172
- Ribas de Pouplana, L., and Schimmel, P. (2001) *Trends Biochem. Sci.* **26**, 591–596
- Traut, T. W. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2944–2948
- Traut, T. W. (1994) *Eur. J. Biochem.* **222**, 9–19
- Bridgham, J. T., Carroll, S. M., and Thornton, J. W. (2006) *Science* **312**, 97–101
- Dean, A. M., and Thornton, J. W. (2007) *Nat. Rev. Genet.* **8**, 675–688
- Benner, S. A., Sassi, S. O., and Gaucher, E. A. (2007) *Adv. Enzymol. Relat. Areas Mol. Biol.* **75**, 9–140
- Gaucher, E. A., Govindarajan, S., and Ganesh, O. K. (2008) *Nature* **451**, 704–707
- Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., and Thornton, J. W. (2007) *Science* **317**, 1544–1548
- Bridgham, J. T., Ortlund, E. A., and Thornton, J. W. (2009) *Nature* **461**, 515–519
- Burbaum, J. J., Starzyk, R. M., and Schimmel, P. (1990) *Proteins Struct. Funct. Genet.* **7**, 99–111
- Fersht, A. R., and Kaethner, M. M. (1976) *Biochemistry* **15**, 3342–3346
- Jakubowski, H., and Goldman, E. (1992) *Microbiol. Mol. Biol. Rev.* **56**, 412–429
- Lincecum, T. L., Jr., Tukalo, M., Yaremchuk, A., Mursinna, R. S., Williams, A. M., Sproat, B. S., Van Den Eynde, W., Link, A., Van Calenbergh, S., Grötl, M., Martinis, S. A., and Cusack, S. (2003) *Mol. Cell* **11**, 951–963
- Mascarenhas, A. P., and Martinis, S. A. (2008) *Biochemistry* **47**, 4808–4816
- Rodin, S. N., and Ohno, S. (1995) *Orig. Life Evol. Biosph.* **25**, 565–589
- Pham, Y., Li, L., Kim, A., Erdogan, O., Weinreb, V., Butterfoss, G. L., Kuhlman, B., and Carter, C. W., Jr. (2007) *Mol. Cell* **25**, 851–862
- Dantas, G., Kuhlman, B., Callender, D., Wong, M., and Baker, D. (2003) *J. Mol. Biol.* **332**, 449–460
- Collaborative Computational Project No. 4 (1994) *Acta Crystallogr.* **D50**, 760–763
- Francklyn, C. S., First, E. A., Perona, J. J., and Hou, Y. M. (2008) *Methods* **44**, 100–118
- SAS Institute (2007) *JMP Statistics and graphics Guide*, SAS Institute, Inc., Cary, NC
- Fersht, A. R., Ashford, J. S., Bruton, C. J., Jakes, R., Koch, G. L., and Hartley, B. S. (1975) *Biochemistry* **14**, 1–4
- Joseph, D. R., and Muench, K. H. (1971) *J. Biol. Chem.* **246**, 7602–7609
- Joseph, D. R., and Muench, K. H. (1971) *J. Biol. Chem.* **246**, 7610–7615
- Hu, H., Boone, A., and Yang, W. (2008) *J. Am. Chem. Soc.* **130**, 14493–14503
- Weinreb, V., Li, L., Campbell, C. L., Kaguni, L. S., and Carter, C. W., Jr. (2009) *Structure* **17**, 952–964
- Liu, Y., and Kuhlman, B. (2006) *Nucleic Acids Res.* **34**, W235–W238
- Dokholyan, N. V., and Shakhnovich, E. I. (2001) *J. Mol. Biol.* **312**, 289–307
- Korczak, B., Le, T., Elowe, S., Datti, A., and Dennis, J. W. (2000) *Glycobiology* **10**, 595–599
- Augustine, J., and Francklyn, C. (1997) *Biochemistry* **36**, 3473–3482
- Schwob, E., and Söll, D. (1993) *EMBO J.* **12**, 5201–5208
- Dubois, D. Y., Blais, S. P., Huot, J. L., and Lapointe, J. (2009) *Biochemistry* **48**, 6012–6021
- Lechler, A., Martin, A., Zuleeg, T., Limmer, S., and Kreutzer, R. (1997) *Nucleic Acids Res.* **25**, 2737–2744
- Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2008) *Nucleic Acids Res.* **36**, D281–D288
- Andreeva, A., Howorth, D., Chandonia, J. M., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2008) *Nucleic Acids Res.* **36**, D419–D425
- Pearl, F. M., Bennett, C. F., Bray, J. E., Harrison, A. P., Martin, N., Shepherd, A., Sillitoe, I., Thornton, J., and Orengo, C. A. (2003) *Nucleic Acids Res.* **31**, 452–455
- Francklyn, C., and Schimmel, P. (1989) *Nature* **337**, 478–481
- Francklyn, C., and Schimmel, P. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8655–8659
- Nixon, A. E., Warren, M. S., and Benkovic, S. J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1069–1073
- Cammer, S., and Carter, C. W., Jr. (2010) *Bioinformatics* **26**, 709–714
- Kapustina, M., Weinreb, V., Li, L., Kuhlman, B., and Carter, C. W., Jr. (2007) *Structure* **15**, 1272–1284
- Deleted in proof
- Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) *Science* **301**, 964–967
- Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) *Science* **292**, 498–500
- Niwa, N., Yamagishi, Y., Murakami, H., and Suga, H. (2009) *Bioorg. Med. Chem. Lett.* **19**, 3892–3894
- Praetorius-Ibba, M., Stange-Thomann, N., Kitabatake, M., Ali, K., Söll, I., Carter, C. W., Jr., Ibba, M., and Söll, D. (2000) *Biochemistry* **39**, 13136–13143
- Bullock, T. L., Uter, N., Nissan, T. A., and Perona, J. J. (2003) *J. Mol. Biol.* **328**, 395–408
- Bullock, T. L., Rodríguez-Hernández, A., Corigliano, E. M., and Perona, J. J. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7428–7433
- Jahn, M., Rogers, M. J., and Söll, D. (1991) *Nature* **352**, 258–260
- Rogers, M. J., Adachi, T., Inokuchi, H., and Söll, D. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3463–3467
- Yaniv, M., Folk, W. R., Berg, P., and Soll, L. (1974) *J. Mol. Biol.* **86**, 245–260
- Wolfson, A. D., Pleiss, J. A., and Uhlenbeck, O. C. (1998) *RNA* **4**, 1019–1023
- Weinreb, V., and Carter, C. W., Jr. (2008) *J. Am. Chem. Soc.* **130**, 1488–1494
- Kumar, R. K., and Yarus, M. (2001) *Biochemistry* **40**, 6998–7004
- Hol, W. G. G., van Duijnen, P. T., and Berendsen, H. J. (1978) *Nature* **273**, 443–446

58. Retailleau, P., Huang, X., Yin, Y., Hu, M., Weinreb, V., Vachette, P., Vonrhein, C., Bricogne, G., Roversi, P., Ilyin, V., and Carter, C. W., Jr. (2003) *J. Mol. Biol.* **325**, 39–63
59. Chuang, W. J., Abeygunawardana, C., Pedersen, P. L., and Mildvan, A. S. (1992) *Biochemistry* **31**, 7915–7921
60. Lockless, S. W., and Ranganathan, R. (1999) *Science* **286**, 295–299
61. Carter, C. W., and Duax, W. L. (2002) *Mol. Cell* **10**, 705–708
62. LéJohn, H. B., Cameron, L. E., Yang, B., MacBeath, G., Barker, D. S., and Williams, S. A. (1994) *J. Biol. Chem.* **269**, 4513–4522
63. LéJohn, H. B., Cameron, L. E., Yang, B., and Rennie, S. L. (1994) *J. Biol. Chem.* **269**, 4523–4531
64. Yang, B., and LéJohn, H. B. (1994) *J. Biol. Chem.* **269**, 4506–4512
65. Williams, T. A., Wolfe, K. H., and Fares, M. A. (2009) *Mol. Biol. Evol.* **26**, 445–450
66. Rodin, A. S., Rodin, S. N., and Carter, C. W., Jr. (2009) *J. Mol. Evol.* **69**, 555–567
67. Kinch, L. N., and Grishin, N. V. (2002) *Curr. Opin. Struct. Biol.* **12**, 400–408