Urzymology: Experimental Access to a Key Transition in the Appearance of Enzymes

Published, JBC Papers in Press, September 10, 2014, DOI 10.1074/jbc.R114.567495
Charles W. Carter, Jr.
From the Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599-7260

Urzymes are catalysts derived from invariant cores of protein superfamilies. Urzymes from both aminoacyl-tRNA synthetase classes possess sophisticated catalytic mechanisms: pre-steady state bursts, significant transition-state stabilization of both amino acid activation, and tRNA acylation. However, they have insufficient specificity to ensure a fully developed genetic code, suggesting that they participated in synthesizing statistical proteins. They represent a robust experimental platform from which to articulate and test hypotheses both about their own ancestors and about how they, in turn, evolved into modern enzymes. They help reshape numerous paradigms from the RNA World hypothesis to protein structure databases and allostery.

Novel Methods and Targets

Urzyme derives from the German/Dutch prefix "Ur," which means primitive or authentic. Woese (1) used the related term "ur-enzyme," which he attributed to K. C. Atwood. As the name suggests, my colleagues and I believe that Urzymes represent legitimate experimental models for very early ancestral enzymes. This belief is based on phylogenetic evidence, their conservation in modern enzymes, and biochemical evidence, their surprisingly high catalytic rate enhancements.

On an evolutionary scale (Fig. 1), Urzymes occupy a gap that is inaccessible to other approaches to evolution. Urzymology has different objectives and methods from ancestral gene reconstruction and directed evolution. During most of biological evolution, fully articulated enzymes assumed new functions via point mutations and short insertions and/or deletions: a process called neofunctionalization (2). Many such events can be recovered from multiple sequence alignments by ancestral gene reconstruction (3–9). Directed evolution uses selection to evoke changes in contemporary enzymes.

My colleagues and I seek to characterize catalysts that are 50–85% smaller and more remote than their contemporary descendants, and are missing entire domains of genetic information. Loss of these substantial modules creates a crucial barrier to accessing early developmental evolution; phylogenetic trees based on multiple sequence alignments representing essentially modern enzymes lose coherence at that stage and do not root in the invariant structural cores (10).

Urzymology uses three-dimensional structural superposition to identify invariant cores. Novel combinations of modular protein engineering, design, biochemistry, and high sensitivity assays allow my colleagues and I then to reconstruct and analyze representations of simpler primordial enzymes that have long been extinct in any form, and whose accessible descendants are much more sophisticated.

Defining evolutionary intermediates is analogous to proving chemical reaction intermediates (11). Intermediate states must be identified and characterized. Then, they must be shown to result from a plausible evolutionary precursor. Finally, they must be shown to give rise to more modern molecular species by similarly plausible evolutionary processes. Biological evolution intermediates are also constrained by using phylogenetic methods to trace ancestries.

This logic extends that articulated by Thornton et al. (12). Urzymes inferred from multiple structure alignments are not true ancestors, but are our best approximations. Given their homology to contemporary enzymes and their robust catalysis, it is very likely that the catalytic cores of these enzymes conserved across all branches of the tree of life point to a quite similar common molecular ancestor. Moreover, irrespective of how closely they resemble true ancestral forms, their activities represent a powerful new tool for studies of contemporary enzyme mechanisms.

The gap between the earliest peptide catalysts and modern enzymes represents an intellectually challenging era that coincides with the emergence of the genetic code, and hence of both genetics and biology. Codon-dependent translation is the nexus between chemical and biological evolution. Amino acid activation and tRNA acylation are necessary and sufficient to translate the genetic code (supplemental Fig. S1). Thus, my colleagues and I have studied Urzymes derived from aminoacyl-tRNA synthetases (aaRS) (13–18). aaRS form two distinct superfamilies with unrelated primary, secondary, and tertiary structures, as well as with significant mechanistic differences (19) (supplemental Figs. S2 and S3).

Fig. 2 illustrates both the notion of and the precedent involved in creating Urzymes. Schwob and Söll (20) nearly produced the first Urzyme by selecting random internal deletions of a suppressor glutaminyl-tRNA synthetase (GlnRS; Class I) for their ability to rescue an amber lacZ mutation. Much of the connective peptide 1 (CP1) insertion and the anticodon-binding domain (ABD) could be removed without eliminating suppression, and a “minimal GlnRS” fusion of the largest CP1 and ABD deletions also restored prototrophy on minimal plates.

* This work was supported, in whole or in part, by National Institutes of Health Grants GM 78227 and GM 90406 through the NIGMS (to C. W. C.). This is the third article in the Thematic Minireview series “Enzyme Evolution.”

2 The abbreviations used are: aaRS, aminoacyl-tRNA synthetase(s); CP1, connecting peptide 1; ABD, anticodon-binding domain; TrpRS, tryptophanyl-tRNA synthetase; TyrRS, tyrosyl-tRNA synthetase; LeuRS, leucyl-tRNA synthetase; HisRS, histidyl-tRNA synthetase; GlnRS, glutaminyl-tRNA synthetase; RO, Rodin-Ohno.

© 2014 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.
The length of this minimal GlnRS (260 residues) is still roughly twice the size of the putative GlnRS Urzyme, which is shown as a ribbon within the context of the full-length GlnRS.

Similar work established the functionality of tRNA acceptor stems (21, 22) and microhelices (23), which could be acylated by intact cognate aaRS. Thus, contemporary aaRS and tRNAs both contain functional subsets that are 50–85% smaller than their full-length relatives. The distinction between these forms and their full-length relatives is both quantitative and qualitative. They have fewer domains and exhibit significant reductions in catalyzed rates.

The radical, directed protein surgery necessary to create Urzymes was motivated by the Rodin-Ohno (RO) hypothesis (24) (supplemental Figs. S2 and S3) that ancestral Class I and Class II aaRS coding sequences were originally complementary strands of the same gene. The only segments of the superfamilies that could be aligned antiparallel as suggested by Rodin and Ohno also turned out to be the invariant cores that position active site residues. The RO hypothesis thus predicted that these cores were intermediates in aaRS evolution, and hence should be catalytically active (18).

Excising Urzymes from full-length enzymes exposes many hydrophobic side chains. These residues must be identified and mutated to restore solubility. Computational methods identified side chains with the greatest newly generated solvent-accessible surface area. Suitable mutations were suggested by the Rosetta protein design program (25). Native TrpRS Urzyme sequences at interfaces with deleted sequences are active when the wild type Urzyme is fused to the anticodon-binding domain (15).

Physical principles motivating the choices that Rosetta makes for these extinct sequences may overlap those induced by selective pressures for stability (26), evoking surrogates for sequence information missing in multiple sequence alignments from living organisms. Thus, protein design extends the study of protein evolution substantially closer to the origin of life.

My colleagues and I are fortunate to have begun by investigating aaRS Urzymes. Urzymes derived from the invariant structural cores of Class I TrpRS (17), LeuRS, and Class II HisRS (16) all have only 15–25% of the total contemporary mass, yet they accelerate both amino acid activation and tRNA aminoacylation proportionately by 100-fold over the uncatalyzed rates (27).

Their strong phylogenetic support (structural invariance) and high catalytic efficiencies afford a robust platform. They point both backward in time to yet more ancient antecedents and forward to the fully developed genetic code (Fig. 1). Thus, their catalytic activities establish a base camp for articulating and testing new and previously inaccessible experimental studies to identify and test intermediates in molecular evolution and allostery (13, 27).

Urzyme Catalytic Activities Satisfy Complementary Tests of Authenticity

Three lines of evidence, pre-steady state burst size, sensitivity to mutation, and substrate binding affinity, reinforce the conclusion that the Urzymes themselves are the authentic sources of the observed catalytic activities. Most unexpected was that the TrpRS, LeuRS, and HisRS Urzymes all exhibit pre-steady

---

3 O. Erdogan and M. Collier, unpublished results.
state bursts comparable in magnitude with the catalyst concentra-

tions, ruling out contamination by tiny amounts of full-
length aaRS (16–18). Bursts established that rate-limiting

product release (28) was a third fundamental link to contempo-

rary enzymes, in addition to accelerating amino acid activation

and tRNA aminoacylation (see supplemental Fig. S1).

Mutational and protein engineering experiments also sup-

port the authenticity of aaRS Urzyme catalytic activities. TrpRS

and HisRS Urzymes, expressed as maltose-binding protein

(MBP) fusions, are activated ~50-fold by tobacco etch virus

protease cleavage. Further, four different HisRS Urzymes dif-

fering in the presence or absence of class-defining signature

Motif 3 and a 6-residue N-terminal extension exhibit catalytic
differences consistent with catalytic contributions of each

module plus a significant \(\sim 1.6 \text{ kcal/mole}\) synergetic inter-

action between them (16). Finally, point mutation of active-site

residues alter catalytic activity by an order of magnitude or

more (16, 17). None of these effects are consistent with activity

from a contaminating catalyst.

Steady-state kinetic parameters afford a third line of evi-
dence for authenticity. Both TrpRS and HisRS Urzymes bind

ATP tightly, but amino acid affinities are 10–100-fold lower

than those of the full-length enzymes. The TrpRS Urzyme tryp-

tophan \(K_m\) is 1–2 mM, 500 times that of intact TrpRS (17). Weak
cognate amino acid binding suggests that discrimination

against similar, non-cognate amino acids is also weakened, as

observed (13, 15, 17).

**aaRS Urzymes: Low Specificity, High Proficiency

Catalysts**

aaRS Urzyme catalytic activities are dramatically higher than

estimates for the uncatalyzed rates (Fig. 3A). Notably, Urzymes

from both classes accelerate both activation and acylation \(10^5–

10^6\)-fold more than necessary to launch ribosome-independent

protein synthesis (Fig. 3B). Further, the two classes achieve sim-

ilar activities with similar masses, consistent with their joint

requirement for protein synthesis. Thus, both classes appear to

have achieved comparable proficiency increments as their

everthing different domain architectures grew comparably in

size.

Urzyme specificities are low. Spectra for Class I LeuRS and

Class II HisRS Urzyme specificities (Fig. 3C) are similar and

complementary. Both Urzymes activate a range of non-cognate

amino acids. Nonetheless, each Urzyme exhibits an ~5-fold

preference for amino acids from its own class. Unknown differ-
MINIREVIEW: Urzymology and the Appearance of Enzymes

ences between aaRS Urzymes and the true ancestral forms may account for some of their promiscuity.

The TrpRS Urzyme provided a unique “molecular knock-out” lacking the entire CP1 and ABD, affording a baseline against which to determine contributions of the two deleted modules. Neither CP1 nor the ABD restored any specificity (15), which results entirely from their energetic coupling.

Using allosteric interactions between genetic modules entirely absent from the Urzymes to enhance specificity resolves challenges (29–33) associated with failure of rational amino acid-binding pocket point mutants to accomplish anything but reducing catalytic activity. Moreover, generating orthogonal synthetase-tRNA pairs appears to require pruning all amino acid-binding residues to alanine and using random mutation to select rebuilt pockets to match the altered substrates consistent with induced-fit mechanisms (34).

These partial results suggest that aaRS Urzymes could not support a canonical 20-amino acid alphabet. Low Urzyme specificity and the fact that Urzymes cannot utilize the tRNA anticodon for recognition form the first experimental basis for the conjecture of Woese (1, 35) that the first coded proteins were statistical ensembles, without unique sequences.

Reduction and Recapitulation

“You have to deeply understand the essence of a product to be able to get rid of the parts that are not essential.” (Jony Ive, quoted in Ref. 36)

Urzymes afford a robust platform for reductionist experiments aimed at characterizing even simpler ancestral protein catalysts that look backward in time (37, 39) and for recapitulating plausible intermediates to test possible evolutionary paths that look forward in time (15) (Fig. 1).

The ability to measure the subtle effects of modules as small as 6–20 amino acids (16) greatly enhances the resolution of modular deconstruction as a tool in protein science. Radical surgery of Class II aaRS afforded evidence that Motif 3, considered essential to catalytic activity because of its interactions with ATP, is dispensable and synergistic even with modular additions elsewhere, including a 6-amino acid N-terminal extension to Motif 1 (15). The catalytic role of Motif 3 may be realized fully only when the Class II insertion domain is present between the Urzyme and Motif 3 (40), or in full-length HisRS.

Further experiments are necessary to map the intermodular synergy.

The unprecedented radical protein surgery that gave rise to the Class I TrpRS and LeuRS Urzymes entailed removing one or more long internal peptides as well as conventional truncation of the C-terminal anticodon-binding domain. Deleting CP1, an internal subdomain, was non-trivial because the two remaining fragments had to be joined together without corrupting the active site. The LeuRS Urzyme entailed removing CP2, in addition to CP1. Removing internal segments was facilitated by the fact that α carbon atoms of their N- and C-terminal residues in full-length aaRS are separated by the length of a peptide bond (13).

RO Hypothesis Redux

The RO hypothesis may have been ignored because it was not obvious how to test it. The relevant objects, ancestral Class I and II aaRS, are so remote that it was hardly evident that the hypothesis could be falsified (48). My colleagues and I articulated and verified bioinformatic (14) and biochemical (16–18, 27) predictions. The balanced, proportionate rate accelerations of both amino acid activation and tRNA aminoacylation by TrpRS and HisRS Urzymes confirmed the prediction that Class I and II segments consistent with antiparallel alignment should be active (27).

Bioinformatic predictions of sense/antisense coding ancestry were tested by excerpting a 94-residue Urgene from ~200 contemporary coding sequences of Class I TrpRS and Class II HisRS. Tyrosyl-tRNA synthetase (TyrrS) and prollyl-tRNA synthetase (ProRS) served as outgroups in rooting the respective trees. Codon middle bases formed base pairs in ~0.34 of all-by-all antiparallel alignments in all four cases, with a standard error of <0.0003, as compared with a well established value of 0.25 for the null hypothesis (supplemental Fig. S4). Middle-base pairing increased in independently reconstructed ancestral sequences for the two trees (supplemental Fig. S5 (14). Middle-base pairing of sense/antisense-related sequences thus

4 C. W. Carter, Jr., and R. Wolfenden, manuscript in preparation.
5 M. Collier, O. Erdogan, and C. W. Carter, Jr., manuscript in preparation.
appears to be a phylogenetic metric that persists far deeper into the past than do metrics for multiple sequence alignments for a single phylogenetic tree.

Their sophistication and tuned catalytic activities argue that aaRS Urzymes had far simpler ancestors. The sense/antisense ancestry of the Urzymes suggests in turn that these simpler ancestors also were encoded on opposite strands of the same gene(s). Phylogenetic evidence and the similar comparisons between two Class IC and two Class IIA aaRS sequence alignments (14) and analysis of amino acid activation by 46-residue ATP-binding sites coded by a designed sense/antisense gene (37) suggest, in turn, that certain properties of these ancestors may be accessible via methods analogous to those of ancestral gene reconstruction (8, 12). See Ref. 39 for additional details.

Urzymology-driven Paradigm Shifts

My colleagues and I expected that experimental studies of ancestral aaRS Urzymes would identify novel perspectives on the origins of translation itself and hence in contemporary molecular biology and biochemistry. However, unexpected new perspectives on phylogenetics/genomics and the origins of protein folding, catalytic activity, specificity, and allostery will also likely shift and enhance comprehensive paradigms in genetics and biophysics.

Codon-dependent Translation

The evolutionary history of the universal genetic code is, in a real sense, that of the two synthetase superfamilies and their cognate tRNAs. Support for the RO hypothesis argues against paradigms holding that the aaRS classes appeared independently, one after the other (49, 50). Urzyme tRNA acylation activity opens to more detailed testing the proposal that early translation used an “operational RNA code” (51) vested only in the tRNA acceptor stem bases, the only parts of tRNA that can be recognized by aaRS Urzymes.

Sense/antisense coding projects further into the past than other metrics (14), implying that catalytic peptides responsible for activating amino acids co-evolved with tRNA from a very primitive state. Thus, contrary to the prevailing RNA World hypothesis, my colleagues and I restated (27) the proposal (52, 53) that genetic coding emerged from mutually catalytic RNA and peptides, using rudimentary stereochemical coding between the two biopolymers.

Phylogenetics/Genomics

Systematic protein structure classifications, SCOP (54) and CATH (55), fail to identify Urzymes of either aaRS class as ancestral forms (10). However, aaRS Urzymes represent plausible ancestors for a wide spectrum of contemporary proteins. The Rossmannoid superfamily (56), the biggest in the proteome (57), includes consensus homologs of Class I aaRS. My colleagues and I argued (58) that the 26 families in the HSP70/actin ATPase superfamily are ancient paralogs of Class II aaRS (pfam: CL0108).

Class II aaRS Urzymes illustrate a distinct, but related problem. One might expect their descendants to include a propor-

FIGURE 4. Structural homology between Class I Urzymes and the TOPRIM domain found in topoisomerases and primases (41). Core domain structures are superimposed in the center. Insertion points corresponding to Class I CP1 and CP2 (42) are shared by various TOPRIM domains, as indicated. The N and C termini of the insertions are indicated by red spheres and correspond closely throughout the domain superfamily. P. horikoshii, Pyrococcus horikoshii; D. radiodurans, Deinococcus radiodurans; E. coli, Escherichia coli.
MINIREVIEW: Urzymology and the Appearance of Enzymes

The most significant puzzle arising thus far from studies of urzymology is that enhanced specificity of contemporary TrpRS, relative to its Urzyme, involves negative epistasis, requiring allosteric energetic coupling between the ABD and the CP1 insertion via a switching element intrinsic to the Urzyme (13, 15). Intramolecular epistasis may have developed while different modules present in contemporary aaRS functioned in trans. Some authors suggest that off-loading functions such as specificity to allosteric effects may be beneficial by making specificity a more robust property (68). The dynamic switching element responsible for this coupling is a widely conserved packing motif (47). How this “protoallosteric” motif (69) functions without the ABD and CP1 modules and how coupling might have emerged before modules became covalently joined remain outstanding questions.

Like knock-out mice, Urzymes are extensive molecular knock-outs that provide unique experimental baselines for helping to answer such questions. Their measurable catalytic rates facilitate multidimensional thermodynamic analysis of modular epistasis in mechanistic enzymology (13, 15).

Protein Design
Recovering Urzymes by deleting non-essential protein masses has adverse effects on stability and solubility. Mutations, identified by Rosetta (60, 61), compensate for these effects. Interfaces between Urzymes and more recently acquired modules can, in principle, also be redesigned. As proficient, relatively nonspecific catalysts, Urzymes can likely be engineered to acylate tRNAs with non-canonical amino acids (34, 70) for industrial purposes.

Urzymes Have Measurable Activities
Most enzyme-catalyzed rates are within the same order of magnitude, irrespective of the uncatalyzed reaction rates (63). Roughly comparable enzyme-catalyzed rates appear to be a requirement for biology. An important implication is that catalytic activities of different enzymes have always been subject to such a constraint. The $10^8$-fold rate accelerations observed for Class I and II aaRS Urzymes (3, 6) suggests that other Urzymes are likely also to have measurable catalytic rates.

Conclusions
In launching urzymology, my colleagues and I took three steps for which there was little, if any, precedent: (i) validating the Rodin-Ohno hypothesis; (ii) testing the evolutionary implication that the most highly conserved portions of enzymes (i.e. their invariant cores) probably have significant catalytic activities; and (iii) using protein design (Rosetta (25)) to compensate for protein mass lost on creating Urzymes, facilitating radical protein surgery. Previously, there was no way to formulate or test hypotheses about how simple, extinct ancestors came to resemble contemporary proteins. Urzymology now offers coherent paradigms that open unprecedented experimental access to mechanisms of very early protein evolution, as well as to novel and effective studies of contemporary mechanistic enzymology.
The first examples of urzymology introduced the ability to use high resolution modular engineering pro-actively, to address previously inaccessible questions. My colleagues and I developed Class I and II aaRS Urzymes to test the Rodin-Ohno hypothesis that ancestral forms of each class descended from opposite strands of the same gene. Unexpected results from that effort will likely change prevailing paradigms in several areas. In validating the hypothesis, my colleagues and I discovered that urzymology, because it connects the earliest genetic coding to the emergence of modern enzymes, affords a powerful enabling technology for studying a key transitional period in the evolution of the genetic code. aaRS Urzymes afford a base camp for probing even more primitive peptide catalysts and for recapitulating subsequent evolutionary steps leading to the emergence of full-length aaRS.

Acknowledgments—I am grateful to numerous laboratory members for their contributions over the past decade and to Richard Wofden for helpful discussions.

REFERENCES

1. Woese, C. R. (1965) On the origin of the genetic code. Proc. Natl. Acad. Sci. U.S.A. 54, 1546–1552
2. Rastogi, S., and Liberles, D. A. (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol. Biol. 5, 28
3. Hanson-Smith, V., Kolaczkowski, B., and Thornton, J. W. (2010) Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. Mol. Biol. Evol. 27, 1988–1999
4. Bridgham, J. T., Ortlund, E. A., and Thornton, J. W. (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461, 515–519
5. Dean, A. M., and Thornton, J. W. (2007) Mechanistic approaches to the study of evolution: the functional synthesis. Nat. Rev. Genet. 8, 675–688
6. Thornton, J. W. (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. Nat. Rev. Genet. 5, 366–375
7. Stackhouse, J., Presnell, S. R., McGehee, G. M., Nambiar, K. P., and Carter, C. W., Jr. (2013) Enhanced amino acid selection in fully evolved Urzymes recapitulating subsequent evolutionary steps leading to the emergence of full-length aaRS.
8. Gaucher, E. A., Govindarajan, S., and Ganesh, O. K. (2008) Palaeotem-
9. Stackhouse, J., Presnell, S. R., McGehee, G. M., Nambiar, K. P., and Carter, C. W., Jr. (2013) Aminoacylating Urzymes for helpful discussions.

MINIREVIEW: Urzymology and the Appearance of Enzymes

10. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code for glutamine. Biochemistry 40, 676–687
11. Fersht, A. R. (1987) Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. Biochemistry 26, 8031–8037
12. Perona, J. I., and Gruic-Sovulj, I. (2014) Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. Top. Curr. Chem. 344, 1–41
13. Corigliano, E. M., and Perona, J. J. (2008) Architectural underpinnings of the genetic code for glutamine. Biochemistry 47, 676–687
14. Bullock, T. L., Rodríguez-Hernández, A., Corigliano, E. M., and Perona, J. I. (2008) A rationally engineered misacylating aminoacyl-tRNA synthetase. Proc. Natl. Acad. Sci. U.S.A. 105, 7428–7433
15. Ibba, M., and Söll, D. (2004) Aminoacyl-tRNAs: setting the limits of the genetic code. Genes Dev. 18, 731–738
16. Praterius-Ibba, M., Stange-Thomann, N., Kitabatake, M., Ali, K., Söll, I., Carter, C. W., Jr., Ibba, M., and Söll, D. (2000) Ancient adaptation of the active site of tryptophanyl-tRNA synthetase for tryptophan binding. Biochemistry 39, 13136–13143
17. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of Escherichia coli. Science 292, 498–500
18. Woese, C. (1998) The universal ancestor. Proc. Natl. Acad. Sci. U.S.A. 95, 6854–6859
19. Issacson, W. (2011) Steve Jobs, p. 364, Simon & Schuster, New York
20. Timenez, M., Williams, T., González-Rivera, A. K., Li, L., Erdogan, O., and Carter, C. W. J. (2014) Did Class 1 and Class 2 aminoacyl-tRNA synthetases descend from genetically complementary, catalytically active ATP-binding motifs? Biophys. J. 106, 675a
21. Kumar, R. K., and Yarus, M. (2001) RNA-catalyzed amino acid activation. Biochemistry 40, 6998–7004
22. Carter, C. W., Jr., Li, L., Weinreb, V., Collier, M., Gonzalez-Rivera, K., Jimenez-Rodriguez, M., Erdogan, O., Kuhlman, B., Ambroggio, X., Williams, T., and Chandrasekharan, S. N. (2014) The Rodin-Ohno hypothesis that two enzyme superfamilies descended from one ancestral gene: an unlikely scenario for the origins of translation that will not be dismissed. Biol. Direct 9, 11
23. Augustine, J., and Franklyn, C. (1997) Design of an active fragment of a
MINIREVIEW: Urzymology and the Appearance of Enzymes

Class II aminoacyl-tRNA synthetase and its significance for synthetase evolution. *Biochemistry* **36**, 3473–3482

41. Aravind, L., Leipe, D. D., and Koonin, E. V. (1998) Toprim: a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**, 4205–4213

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

43. Gong, L., Suchard, M. A., and Bloom, J. D. (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631

44. Nasrallah, C. A., and Huelsenbeck, J. P. (2013) A phylogenetic model for the detection of epistatic interactions. *Mol. Biol. Evol.* **30**, 2197–2208

45. Covert, A. W., Jr., and Kraut, J. (1974) A proposed model for interaction of tryptophanyl-tRNA synthetase. *Structure in Proteins* **7**, 99–111

46. Popper, K. R. (1959) *The Logic of Scientific Discovery*, pp. 124–130, Routledge, Oxford, United Kingdom

47. Cammer, S., and Carter, C. W., Jr. (2010) Six Rossmannoid folds, including the Class I aminoacyl-tRNA synthetases, share a partial core with the anticodon-binding domain of a Class II aminoacyl-tRNA Synthetase. *Bioinformatics* **26**, 709–714

48. Klipcan, L., and Safro, M. (2004) Molecular evolution: aminoacyl-tRNA synthetases on the loose. *Curr. Biol.* **9**, R842–R844

49. Weiner, A. M. (1999) Molecular evolution: aminoacyl-tRNA synthetases on the loose. *Curr. Biol.* **9**, R842–R844

50. Schimmel, P., Giegé, R., Moras, D., and Yokoyama, S. (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8763–8768

51. Carter, C. W., Jr. (1975) Cradles for molecular evolution. *New Sci.* **65**, 784–787

52. Carter, C. W., Jr., and Kraut, J. (1974) A proposed model for interaction of polypeptides with RNA. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 283–287

53. Andreeva, A., Howorth, D., Chandonia, J. M., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2008) Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* **36**, D419–D425

54. Pearl, F. M., Bennett, C. F., Bray, J. E., Harrison, A. P., Martin, N., Sheikh, A., Sillitoe, I., Thornton, J., and Orengo, C. A. (2003) The CATH database: an extended protein family resource for structural and functional genomics. *Nucleic Acids Res.* **31**, 452–455

55. Goldstein, R. (2008) The structure of protein evolution and the evolution of protein structure. *Curr. Opin. Struct. Biol.* **18**, 170–177

56. Stockbridge, R. B., Lewis, C. A., Jr., Yuan, Y., and Wolfenden, R. (2010) Impact of temperature on the time required for the establishment of priomordial biochemistry, and for the evolution of enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 22102–22105

57. Shakhnovich, B. E., Dokholyan, N. V., DeLisi, C., and Shakhnovich, E. I. (2003) Functional fingerprints of folds: evidence for correlated structure-function evolution. *J. Mol. Biol.* **326**, 1–9

58. 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

59. Artymiuk, P. J., Rice, D. W., Poirrette, A. R., and Willet, P. (1994) A tale of two synthetases. *Nat. Struct. Biol.* **1**, 758–760

60. 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

61. Wolfenden, R. (2011) Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Ann. Rev. Biochem.* **80**, 645–667

62. Stockbridge, R. B., Lewis, C. A., Jr., Yuan, Y., and Wolfenden, R. (2010) Impact of temperature on the time required for the establishment of priomordial biochemistry, and for the evolution of enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 22102–22105

63. Schimmel, P., Giegé, R., Moras, D., and Yokoyama, S. (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8763–8768

64. Weinreb, V., Li, L., Kaguni, L. S., and Carter, C. W., Jr. (1990) Mg²⁺-assisted catalysis by *B. stearothermophilus* TrpRS is promoted by allosteric effects. *Structure* **17**, 952–964

65. Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code. *Ann. Rev. Biochem.* **79**, 431–444

66. Schroeder, G. K., and Wolfenden, R. (2007) The rate enhancement produced by the ribosome: an improved model. *Biochemistry* **46**, 4037–4044

67. Stockbridge, R. B., and Wolfenden, R. (2009) The intrinsic reactivity of ATP and the catalytic efficiencies of kinases acting on glucose, N-acetylglucosamine, and homoserine. *J. Biol. Chem.* **284**, 22747–22757