Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 327–328, January 1999

Commentary

Genetic code origins: Experiments confirm phylogenetic predictions and may explain a puzzle

Paul Schimmel and Lluis Ribas de Pouplana*

The Skaggs Institute for Chemical Biology, The Scripps Research Institute, Beckman Center, 10550 North Torrey Pines Road, La Jolla, CA 92037

For many years the question of the origin of the genetic code has both fascinated and frustrated serious investigators. On the one hand, the elegant simplicity and universal appearance of the code fascinate them. What could be more fundamental than the code to understanding life's origins? On the other hand, they are frustrated by the obvious challenge of figuring out how to go back in time and reconstruct early events associated with the code. In particular, what kind of contemporary analyses and experiments can be done to reconstruct past events related to the genetic code? A partial answer to the last question is now coming from phylogenetic analyses of the information generated by the many genome projects. But the missing connection has been experiments that test specific predictions of those analyses. Now, in this issue of the *Proceedings*, Ibba *et al.* (1) present experiments that test phylogenetic predictions and, in addition, may explain a puzzle.

Transfer RNAs and tRNAs synthetases are the center of attention for phylogenetic studies. The reason is that the code is established in aminoacylation reactions catalyzed by aminoacyl tRNA synthetases. There is (with some exceptions) one tRNA synthetase for each amino acid and, because of the degeneracy of the code, one or more tRNAs for each amino acid. In the aminoacylation reactions, each amino acid is joined to the tRNA that harbors the anticodon triplet of the code for that amino acid. In this way, the algorithm that relates amino acids to nucleotide triplets is universally determined. Thus, the question of the origin of the code has to deal at some point with tRNAs and tRNA synthetases.

A number of considerations led to suggestions that tRNAs preceded their synthetases in evolution (2–7). This order of appearance fits with the concept of an early RNA world that preceded the theater of proteins (8–14). It also fits with the experimental demonstration of the capacity of RNA to catalyze aminoacylation (15, 16). But until recently, no phylogenetic analyses were possible that explicitly tested the order of appearance of tRNAs and their synthetases. This situation changed, however, because of another finding that, in turn, led to a phylogenetic study that specifically concluded that at least one tRNA appeared before a cognate synthetase (17).

The basic idea is that the tRNA synthetases are divided into two distinct classes (class I and class II) of 10 enzymes each (18, 19). The two classes differ in three ways: the architecture of the active site region; the tRNA hydroxyl group (2′- or 3′-) to which the enzyme initially joins the amino acid; and the side (major or minor groove) of the tRNA amino acid acceptor stem to which the enzyme binds. No evidence supports the existence of a common ancestor for the two groups. Most typically, the class to which an enzyme belongs is fixed in evolution through all three kingdoms——bacteria, archae, and eukarya. But lysyl-tRNA synthetase is an exception. In most organisms it is a class II enzyme. However, in bacterial spirochetes such as *Borrelia burgdorferi* and α-proteobacteria such as *Rickettsia prowazekii*, and in many Euryarchaeota (e.g., *Methanococcus jannaschii, Methanococcus maripaludis, Archaeoglobus fulgidus*, among others) and Crenarchaeota, LysRS has the architecture of a class I enzyme (1, 20, 21). In no case is there evidence for more than one lysyl-tRNA synthetase in each of these organisms. This situation made possible a phylogenetic analysis to determine whether the two types of lysyl-tRNA synthetases preceded the appearance of $tRNA^{Lys}$ (17).

Given that the two classes of enzymes approach the tRNA acceptor stem from opposite sides, and given that they initially attach the amino acid to different ribose-hydroxyl groups, two types (clearly distinct sequences) of lysine tRNAs were imagined to accompany the two kinds of lysyl-tRNA synthetases. These two types would be designed for interactions with opposite sides of the acceptor helix and to facilitate initial amino acid attachment to distinct hydroxyl groups. However, the division of lysine tRNAs into two types that mirrored the two types of lysine enzymes was not seen. The lysine tRNAs inferred to be charged by class I enzymes were closely similar to those charged by class II enzymes. Thus, the two types of lysyl-tRNA synthetases adapted to the same tRNA. Thus, at least one lysyl-tRNA synthetase appeared after $tRNA^{Lys}$ was established (17).

Can charging of the same $tRNA^{Lys}$ with members of each of the two distinct classes of lysyl-tRNA synthetases be demonstrated *in vivo* and *in vitro*? And how do you explain the puzzling displacement in some organisms of a lysyl-tRNA synthetase with its opposite type? These are the questions addressed by Ibba *et al*. (1).

For the *in vivo* experiments, they take advantage of a strain of *Escherichia coli* that is deficient in lysyl-tRNA synthetase (22). As a result of this deficiency, the cells cannot grow unless provided with an active lysyl-tRNA synthetase. *E. coli* and most bacterial organisms encode class II lysyl-tRNA synthetases in their genomes. However, expression of the gene for the class I spirochete *B. burgdorferi* lysyl-tRNA synthetase rescues growth. This result is especially significant because mischarging of lysine onto other tRNAs by the *B. burgdorferi* enzyme is expected to result in lethality (23, 24). Thus, a class I enzyme specifically charges (in *E. coli*) the same $tRNA^{Lys}$ as charged by the class II *E. coli* lysyl-tRNA synthetase. Some *in vitro* experiments support this conclusion.

Further experiments *in vivo* show that expression of the plasmid-borne class I Euryarchaeota *M. maripaludis* lysyl-tRNA synthetase also rescues growth. In this instance, growth rescue is not as robust and requires lysine supplementation of the growth media. (The reason for less robust growth is not explored; it could be caused by poor expression or the instability of the *M. maripaludis* enzyme in *E. coli* and, additionally or alternatively, by a low rate of charging of *E. coli* $tRNA^{Lys}$.) Experiments *in vitro* support the conclusion that the *M. maripaludis* enzyme can charge *E. coli* $tRNA^{Lys}$ (although charging efficiency is reduced).

These biochemical experiments nicely confirm the phylogenetic analyses that inferred that at least one of the two types of lysyl-tRNA synthetases appeared after the establishment of $tRNA^{Lys}$ (17). Ibba *et al.* (1) then go on to deal with the question

The companion to this Commentary is published on page 418.

*To whom reprint requests should be addressed. e-mail: schimmel@scripps.edu.

PNAS is available online at www.pnas.org.

of how one synthetase displaced the other in certain organisms. Here they take up the question of specific nucleotides that are required for charging of $tRNA^{Lys}$——an area where much is already known. The main point they establish is that a specific acceptor stem base pair—a G2:U71 wobble pair at the second position of the helix—blocks charging by the class II enzyme (Fig. 1). This base pair is found in the spirocytes *Treponema pallidum* and *B. burgdoferi* that contain a class I enzyme. In contrast, Ibba *et al.* (1) show that the class II *E. coli* enzyme can accept either G2-C71 (found in organisms containing the class II enzyme) or G2-U71. Thus, the presence of a specific base pair at the end of the acceptor stem correlates with the presence of the less common class I enzyme in *B. burgdoferi*.

Although not explicitly discussed by Ibba *et al.* (1), the variation in the 2:71 base pair alone cannot explain the displacement of one type of lysyl-tRNA synthetase with another. There are many examples of organisms where lysine tRNAs have a G2:C71 base pair, regardless of whether the homologous LysRS is class I or II. These examples include the class I LysRS-containing archaebacterial *M. jannaschii, M. maripaludis, Methanobacterium thermoautotrophicum,* and *A. fulgidus,* and the class II-containing bacterial organisms *E. coli* and *Haemophilus influenzae* and eukaryotes such as *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, and humans. Similarly, the class I LysRS of *R. prowacekii* and the class II LysRS of *Staphylococcus aureus* and *Bacillus subtilis* charge lysine tRNAs that have an A2:U71 pair. So, class-specific discrimination of the 2:71 base pair is not a general phenomenon. Instead, it is restricted to the spirocytes and *E. coli* example chosen by Ibba *et al.* (1). Still, the general principle may be valid. That is, blocking nucleotide determinants (at some position in the tRNA structure) may force out one synthetase and replace it with another.

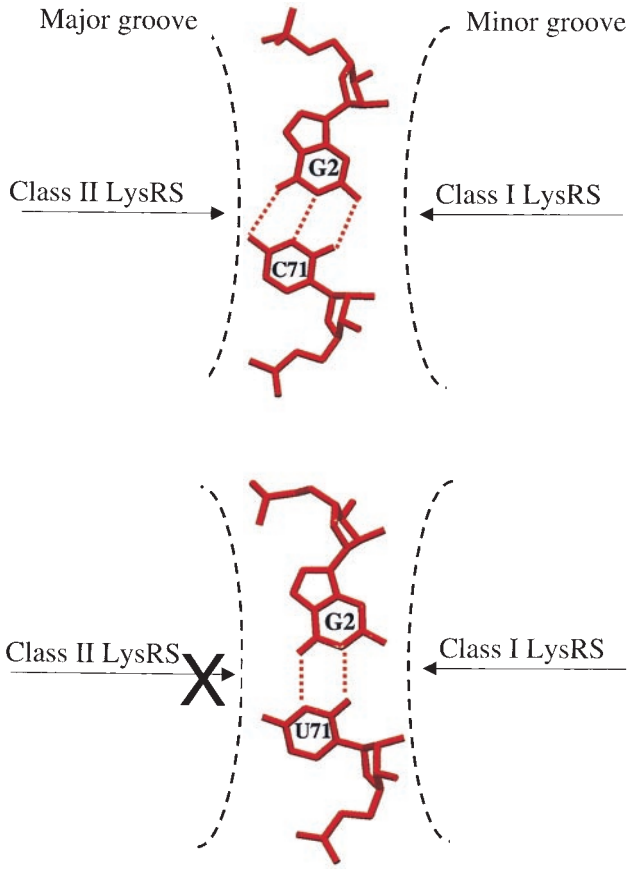

FIG. 1. Schematic depiction of the base pair near the end of the acceptor stem of a lysine tRNA acting as a blocking determinant for one enzyme but not another. The class I lysyl-tRNA synthetase binds to the minor groove side of the helix, whereas the class II enzyme approaches from the major groove side. The class II lysyl-tRNA synthetase of *E. coli* accommodates major groove determinants encoded by G:C, but is blocked by the presence of a G:U pair. In contrast, the class I lysine enzyme of the spirochete *B. burgdorferi* accepts the minor groove determinants of G:C or G:U. The appearance of blocking determinants for a class II lysyl-tRNA synthetase could explain its displacement by its class I counterpart.

When did the substitution of one synthetase for another occur? All four archaebacterial genomes determined to date encode a class I LysRS. All eukaryotes characterized so far have a class II LysRS. The bacterial kingdom has examples of either class, although no organism is yet known to contain one from each class in the same organism. (*E. coli* has two genes for LysRS—both are class II.) So, multiple lateral gene transfer from archaebacteria to certain bacteria could account for the presence of class II LysRS in bacterial organism such as *T. pallidum, B. burgdorferi*, and *R. prowacekii*. But the remains of the displaced genes have not been found. Thus, the extant distribution of the two classes of LysRS could have been established at an early evolutionary stage, through lateral transfer events that occurred before the separation of the three main branches of the tree of life (cf. ref. 25).

1. Ibba, M., Losey, H. C., Kawarabayasi, Y., Kikuchi, H., Bunjun, S. & Söll, D. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 418–423.
2. Woese, C. (1967) *The Genetic Code* (Harper and Row, New York).
3. Weiner, A. M. & Maizels, N. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7383–7387.
4. de Duve, C. (1991) *Blueprint for a Cell: The Nature and Origin of Life* (Patterson Publishers, Carolina Biological Supply Company, Burlington, NC).
5. Maizels, N. & Weiner, A. M. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 577–602.
6. Weiner, A. M. & Maizels, N. (1994) *Curr. Biol.* **4,** 560–563.
7. Nagel, G. M. & Doolittle, R. F. (1995) *J. Mol. Evol.* **40,** 487–498.
8. Cech, T. R. & Bass, B. L. (1986) *Annu. Rev. Biochem.* **55,** 599–629.
9. Altman, S., Baer, M. F., Bartkiewicz, M., Gold, H., Guerrier-Takada, C., Kirsebom, L. A., Lumelsky, N. & Peck, K. (1989) *Gene* **82,** 63–64.
10. Pace, N. R. & Smith, D. (1990) *J. Biol. Chem.* **265,** 3587–3590.
11. Robertson, D. L. & Joyce, G. F. (1990) *Nature (London)* **344,** 467–468.
12. Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **346,** 818–822.
13. Tuerk, C. & Gold, L. (1990) *Science* **249,** 505–510.
14. Noller, H. F., Hoffarth, V. & Zimniak, L. (1992) *Science* **256,** 1416–1419.
15. Piccirilli, J. A., McConnell, T. S., Zaug, A. J., Noller, H. F. & Cech, T. R. (1992) *Science* **256,** 1420–1424.
16. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. (1995) *Science* **267,** 643–647.
17. Ribas de Pouplana, L., Turner, R. J., Steer, B. & Schimmel, P. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 11295–11300.
18. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) *Nature (London)* **347,** 203–206.
19. Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N. & Leberman, R. (1990) *Nature (London)* **347,** 249–255.
20. Ibba, M., Morgan, S., Curnow, A. W., Pridmore, D. R., Vothknecht, U. C., Gardner, W., Lin, W., Woese, C. R. & Soll, D. (1997) *Science* **278,** 1119–1122.
21. Ibba, M., Bono, J. L., Rosa, P. A. & Söll, D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 14383–14388.
22. Chen, J., Brevet, A., Lapadat-Tapolsky, M., Blanquet, S. & Plateau, P. (1994) *J. Bacteriol.* **176,** 2699–2705.
23. Bedouelle, H., Guez, V., Vidal-Cros, A. & Hermann, M. (1990) *J. Bacteriol.* **172,** 3940–3945.
24. Vidal-Cros, A. & Bedouelle, H. (1992) *J. Mol. Biol.* **223,** 801–810.
25. Woese, C. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6854–6859.