Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 4521–4522, May 1996

Commentary

Origin of genetic code: A needle in the haystack of tRNA sequences

Paul Schimmel

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Because the amino acid–trinucleotide algorithm of the genetic code is established by the specific aminoacylations of tRNAs, the sequences and structure of tRNAs have long been investigated with the idea of learning about the possible origin of the code (1). The idea is that elements of the code might appear in parts of the tRNA structure other than the anticodon. These parts might represent primordial components of the tRNA molecule, which possibly served as structures associated with the aminoacylations of particular amino acids (2, 3). The paper by Rodin *et al.* (4) in the current issue of the *Proceedings* points out a previously unrecognized relationship between tRNA sequences that could be relevant to the origin of the code.

Investigations of this sort are inherently difficult. On the one hand, more than 2000 tRNA sequences have been collectively determined for specific eubacterial, archaebacterial, and lower and higher eukaryote organisms (5). This large data base of sequences provides ample material for analyses which attempt to find vestiges of the genetic code in parts of the tRNA structure outside of the anticodon. On the other hand, with only four nucleotides, the random chance of a specific sequence relationship is far higher with tRNAs than with proteins made up of 20 amino acids. The possibilities for random "hits" increase considerably the background "noise" that must be discounted.

The other, related problem is that tRNAs are ancient molecules, probably arising more than one billion years ago with the earliest life forms. Consequently, they have undergone many mutational variations and changes in evolution. With only four nucleotides and a long evolutionary history, searching for vestiges of the genetic code within the tRNA structure is like searching for a small needle in a big haystack.

The main focus of the Rodin *et al.* (4) and other analyses is on tRNA acceptor stems. The tRNA molecule has a cloverleaf secondary structure with four "arms" known as the dihydrouridine stem-loop, the anticodon stem-loop, the TψC stem-loop, and the acceptor stem which terminates at the 3′ end in the single-stranded $N^{73}CCA^{76}$ tetranucleotide common to all tRNAs. The amino acid attachment site (at A^{76}) is at one end of the cloverleaf, and the anticodon trinucleotide of the genetic code is at the other (Fig. 1). These two parts of the molecule are also at opposite ends (separated by *ca.* 75 Å) of the three-dimensional L-shaped tRNA structure, which forms by condensing the four arms of the cloverleaf into two major domains (6). These domains are the acceptor–TΨC minihelix and the anticodon–dihydrouridine stem-biloop.

In the L-shaped structure, the acceptor stem and amino acid attachment site are segregated into a separate domain (minihelix) distinct from the domain containing the anticodon trinucleotide. This segregation can be viewed as functional as well as structural, because acceptor stems alone are substrates for specific aminoacylations by many of the aminoacyl tRNA synthetases (7–9). These aminoacylations constitute an operational RNA code for amino acids whereby sequences and structures in acceptor stems correspond to specific amino acids (10). In addition, when bound to the ribosome, the tRNA molecule segregates interactions of its two domains with separate, distinct ribosomal RNAs (11). The segregation of these binding interactions further emphasizes the functional modularity of the tRNA structure, which separates critical signals for aminoacylation and the amino acid attachment site from the template reading head of the code.

The acceptor stem-containing minihelix has long been thought to predate the full tRNA structure. The minihelix was proposed as a tag required on ancient RNA genomes for their replication (12, 13). It serves the same role on many contemporary genomes, ranging from the bacteriophage Qβ RNA to plant viral RNAs to the *Neurospora crassa* Mauriceville retroplasmid (13, 14). Schemes for the development of the contemporary tRNA structure include, among others, duplications and rearrangements of the basic minihelix hairpin so that the final tRNA structure is visualized, in one way or another, as comprised of two domains arising from a single minihelix (4, 15, 16). In these schemes, the anticodon trinucleotide of the code is seen as arising from specific nucleotides in the acceptor stem, such as a subset of the first few base pairs of the stem. As it turns out, acceptor stem signals for aminoacylation are generally confined to the first four or five base pairs, in addition to the N^{73} "discriminator" base.

With this scenario for the development of the tRNA molecule from a minihelix, sequence analyses can concentrate on identifying acceptor stem nucleotides that are related in some way to one or more of the anticodon trinucleotides or, equivalently, to one or more complementary codon bases in the acceptor stem duplex. The relatively high frequency of codons for alanine, glycine, valine, and aspartate in the acceptor stems of their respective tRNAs was noted by Möller and Janssen (15, 17). This relationship is particularly prominent for tRNAs from chloroplasts and from the kingdoms of eubacteria and archaebacteria. Although these frequencies are high relative to expectations based on the random occurrences of particular trinucleotide sequences calculated with a probability of 1/4 assigned for each base, there may be an element of chance because alanine and glycine tRNAs in particular are G,C-rich and the cognate amino acids have codons that are G,C-rich such as GGC (alanine) and GCC (glycine).

Rodin *et al.* (4) approached the problem from a different perspective which is rooted in their interest in the possible significance of the sequence complementarity of opposing strands of RNA duplexes for the evolution and development of the genetic code. The approach was to look for relationships in tRNA sequences which suggested that those tRNAs with complementary anticodons, for example, also had some kind of complementarity of their acceptor stems. Such relationships could support the hypothesis that one or more anticodon nucleotide was historically related to an acceptor stem nucleotide needed for aminoacylation.

Approximately 1300 tRNA sequences were examined and consensus acceptor stems were constructed for tRNAs corresponding to each of the 20 amino acids. The tRNAs were then organized into all 32 pairs that have complementary anticodons. Of these 32 pairs, 29 also had a complementary relationship at the second position in the acceptor stem.

For example, four of the six-fold degenerate leucine codons are CUN, where N is any of the four bases. The anticodons which read CUN are N′AG, where N′ is complementary to N by Watson–Crick or wobble pairing. One $tRNA^{Leu}$ has the anticodon AAG that is complementary to a $tRNA^{Lys}$ with a

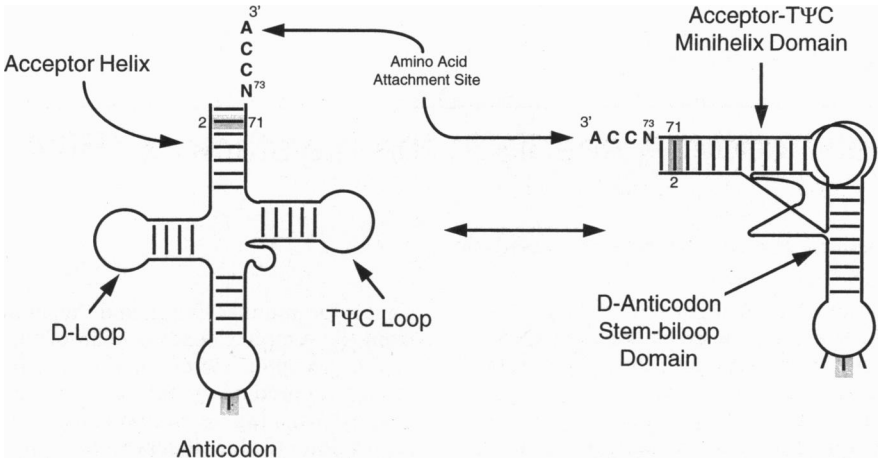

FIG. 1. Schematic diagram of tRNA cloverleaf (*Left*) and of L-shaped three-dimensional structure comprised of two domains (*Right*). The 2–71 bp in the acceptor helix and the second base of the anticodon are shaded. (Illustration provided by Dr. Barry Henderson.)

CUU anticodon. The second position of the acceptor stem of this $tRNA^{Leu}$ is a "G," while this position is the complementary "C" in the $tRNA^{Lys}$ counterpart. Another $tRNA^{Leu}$ isoacceptor has a UAG anticodon that is complementary to a CUA anticodon of a $tRNA^{Gln}$. That $tRNA^{Leu}$ has a "C" at the second position of the acceptor stem, while the particular $tRNA^{Gln}$ has a "G." And so on for the other tRNAs which have complementary anticodons.

These relationships are noteworthy. Rodin *et al.* (4) postulate that the middle (second) position of the anticodon triplet originated from the second position of the historical minihelix. In their scheme for the evolution of tRNAs, opposite strands of RNA duplexes encoded distinct tRNAs. These duplexes, in turn, originated from minihelix-like molecules which replicated and combined to give the two complementary strands that became full tRNAs. The idea is that tRNAs with complementary anticodons were encoded in pairs by opposite strands of an RNA helix. Each strand had an anticodon triplet that originated from an acceptor stem triplet according to the scheme of conversion of a minihelix to a full tRNA (*vide supra*). The second and other acceptor stem nucleotides might change with time in this genome. However, the requirement for complementarity of opposing strands would mean that, as long as the duplex encoded pairs of tRNAs, those with complementary anticodons would also have complementary acceptor stems. The vestige of such a complementary relationship is seen in contemporary tRNAs at the second position of the acceptor stem.

If the second position of the anticodon originated from the second position of the acceptor stem, then the first and third positions of the anticodon presumably would have arisen from the first and third nucleotides of the stem. In that case, the same complementarity relationships should show up for these nucleotides as well. According to Rodin *et al.* (4), the same complementarity relationship is observed if the third position of the acceptor stem is used, although this correlation is less robust. In the case of the first position of the stem, this nucleotide is usually at "G" (possibly for structural reasons to stabilize the end of the acceptor helix) and corresponds to the degenerate wobble position of the anticodon. For these reasons, no complementarity relationship is expected at the first position, and none is observed.

Has a needle in the haystack of tRNA sequences been found? The answer can only be given in statistical, probabilistic terms. Rodin *et al.* (4) point out that, based on random chance, the probability is roughly 1×10^{-6} for having 29 or more out of 32 tRNAs with complementary anticodons also happening to have complementary bases at the second positions of their respective acceptor stems. Thus, the likelihood that this relationship is a random fluke seems low, but not impossible. Meanwhile, the search for the needle goes on.

1. Eigen, M., Lindemann, B. F., Tietze, M., Winkler-Oswatitsch, R., Dress, A. & von Haeseler, A. (1989) *Science* **244,** 673–679.
2. de Duve, C. (1988) *Nature (London)* **333,** 117–118.
3. de Duve, C. (1991) *Blueprint for a Cell: The Nature and Origin of Life* (Neil Patterson/Carolina Biological Supply, Burlington, NC), pp. 173–198.
4. Rodin, S., Rodin, A. & Ohno, S. (1996) *Proc. Natl. Acad. Sci. USA* **93,** in press.
5. Steinberg, S., Misch, A. & Sprinzl, M. (1993) *Nucleic Acids Res.* **21,** 3011–3015.
6. Rich, A. & RajBhandary, U. L. (1976) *Annu. Rev. Biochem.* **45,** 805–860.
7. Frugier, M., Florentz, C. & Giegé, R. (1994) *EMBO J.* **13,** 2218–2226.
8. Schimmel, P. & Ribas de Pouplana, L. (1995) *Cell* **81,** 983–986.
9. Hou, Y.-M., Sterner, T. & Bhalla, R. (1995) *RNA* **1,** 707–713.
10. Schimmel, P., Giegé, R., Moras, D. & Yokoyama, S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8763–8768.
11. Noller, H. F. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 137–156.
12. Weiner, A. M. & Maizels, N. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7383–7387.
13. Maizels, N. & Weiner, A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6729–6734.
14. Wang, H. & Lambowitz, A. M. (1993) *Cell* **75,** 1071–1081.
15. Möller, W. & Janssen, G. M. C. (1992) *J. Mol. Evol.* **34,** 471–477.
16. Dick, T. P. & Schamel, W. W. A. (1995) *J. Mol. Evol.* **41,** 1–9.
17. Möller, W. & Janssen, G. M. (1990) *Biochemistry* **72,** 361–368.