Commentary

Diverse RNA substrates for aminoacylation: Clues to origins?

Paul Schimmel* and Rebecca Alexander

The Skaggs Institute for Chemical Biology, The Scripps Research Institute, Beckman Center, 10560 North Torrey Pines Road, La Jolla, CA 92037

The aminoacylation of RNA structures generally is considered the starting point for the emergence of the theater of proteins from the RNA world. Once an aminoacyl ester linkage is established with an RNA acceptor, peptide bond formation is thermodynamically favored. In contemporary protein biosynthesis, aminoacylation is directed to the 3′-end of the tRNA structure, which, in turn, brings the activated amino acid to the ribosomes for addition to the growing polypeptide chain. But during the last 9 years a variety of artificial RNA substrates for aminoacylation has been reported (1–8). And in this issue of the *Proceedings*, the list of examples is expanded in a significant way. Felden and Giegé (9) describe for the first time an aminoacylation system based on a circular RNA substrate. The results obtained with this novel substrate add further support to the idea that the contemporary system of tRNA aminoacylation could have grown out of early aminoacylation systems that used diverse RNA oligonucleotide substrates. Interactions between some of these aminoacylated substrates could have led to a primitive system of peptide synthesis.

All of these substrates for aminoacylation are based, in one way or another, on one of the two domains within the tRNA structure. The amino acid attachment site is at the end of a long helix (the acceptor-TψC minihelix) that terminates at the 3′-end in the universal sequence NCCA^{OH} (Fig. 1, *Top*). The second domain of the tRNA structure is at right angles (in three dimensions) and consists of a 10-bp helix joined to a hairpin loop that harbors the anticodon triplet of the genetic code. The resulting two-domain L-shaped structure places the amino acid attachment site of the minihelix about 75 Å from the anticodon triplet.

Significantly, the anticodon itself is not used in every case by the cognate tRNA synthetase to identify the correct tRNA for its amino acid. Two prominent examples are alanyl- and seryl-tRNA synthetases, where the respective enzyme makes no contact whatsoever with the anticodon triplet of the cognate tRNA (10–12). Instead, the minihelix domain contains critical determinants for aminoacylation. These minihelix determinants exist even in those tRNAs where the cognate synthetase makes an important contact with the anticodon triplet (2, 4–6, 13–15). The most rigorous proof of the "information content" in the minihelix domain was the direct demonstration that, for alanyl-tRNA synthetase, the minihelix in isolation was a robust substrate. The critical determinant for aminoacylation of tRNA^{Ala} is a G3:U70 base pair in the acceptor helix (16, 17). Transfer of this base pair into other tRNAs conferred alanine acceptance on them. The same properties can be demonstrated for the minihelix—that is, charging depends on the G3:U70 pair and transfer of the pair into other minihelices converted them to alanine acceptors (1). This result showed clearly that the relationship between alanine and the triplets of the genetic code was indirect.

Since this demonstration of minihelix aminoacylation with alanine, a total of 10 different amino acids have been enzymatically charged to minihelix-like substrates. Some of these systems are more robust than others, but in all cases charging is sequence specific. Moreover, nucleotide substitutions in an isolated minihelix that affect charging efficiency also affect charging when placed in the minihelix domain of the corresponding full tRNA. Thus, nucleotide determinants for specific aminoacylation of minihelices have been retained in the contemporary tRNA. Because the minihelix domain is thought to be the more ancient part of the tRNA structure (18–20), aminoacylation of minihelices is imagined to reflect an earlier system of charging. Eventually, the minihelix domain was combined with a second, newer domain that brought in a template reading head containing the anticodon triplets of the genetic code (21, 22).

The contemporary minihelices are not likely identical to the earliest substrates. Rather, other structures—precursors to the minihelix—were probably prominent. The list of laboratory examples of minihelix-like structures and pieces of minihelices has grown steadily, and these examples illustrate the diverse possibilities for early aminoacylation systems that could have developed into the modern minihelix motif. Demonstrated substrates include: RNA duplexes of various lengths; short hairpin oligonucleotides made up of the first 7 bp of the minihelices (designated as microhelices); hairpin duplexes with only 4 bp that are stabilized by a specific tetraloop motif; and a pseudoknot that is created by the hydrogen bonding of nucleotides in a hairpin loop to sequences that are distal to the stem of the hairpin (Fig. 1, *Middle*). In the latest example, Felden and Giegé (9) design and synthesize a knotted RNA circle. The design is a clever modification of a pseudoknot motif found near the 3′-end of brome mosaic virus RNA (23, 24). The heart of the modification includes the introduction of a single-stranded circle that rearranges into a minihelix-like pseudoknot upon hybridization with a short linear fragment. The hybridized linear fragment is charged specifically with histidine (Fig. 1, *Bottom*).

The key to success in all of these examples lies in the localization of nucleotide determinants for aminoacylation to the region near the amino acid attachment site. Typically, one or more of the first 4 bp of the helix and specific 2′-hydroxyl groups play an important role in determining charging efficiency and specificity, as does G3:U70 for tRNA^{Ala} (7). In addition, all substrates end in the universal NCCA^{OH}, where "N" is the so-called discriminator base. [This nucleotide is usually an important determinant of aminoacylation efficiency. Its specific identity originally was proposed to correlate with the type of amino acid (aliphatic, hydrophilic, aromatic, etc.) that was attached to the associated tRNA (25).] These specific minihelix sequences/structures constitute an operational RNA code for amino acids that is distinct from the genetic code. The idea, therefore, is to present each of these determinants in such a way that the synthetase has unobstructed access to them. In the latest study, the knotted circle was designed so that nucleotides near the amino acid attachment site were directly accessible to histidyl-tRNA synthetase. Specific nucleotide substitutions in the knotted circle were made at locations known to be important for charging with
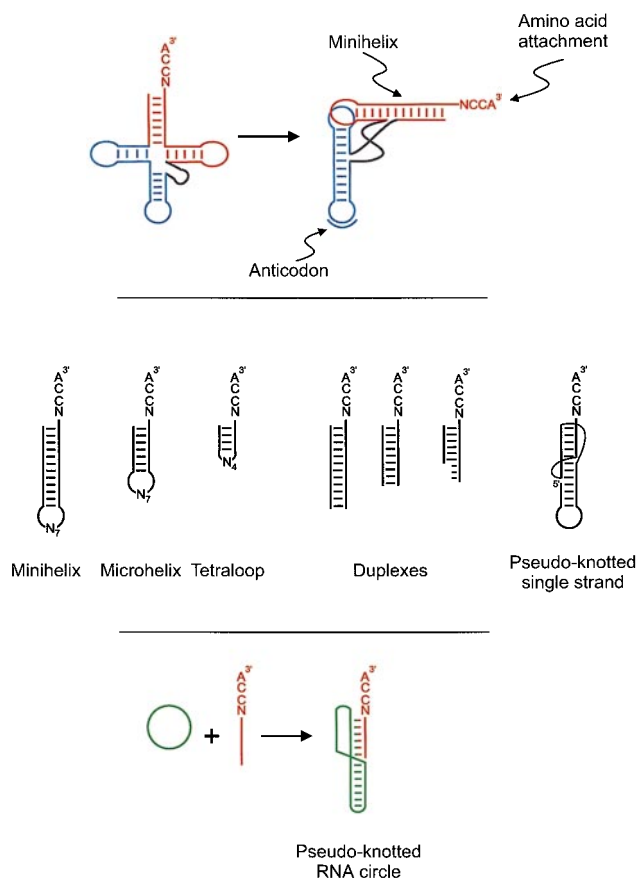
*To whom reprint requests should be addressed.

FIG. 1. The diversity of RNA oligonucleotide substrates for aminoacylation. (*Top*) The tRNA cloverleaf secondary structure (*Left*) that folds into an L-shape (*Right*). The minihelix domain is shaded in red, and the second domain containing the template reading head is shown in blue. (*Middle*) The variety of oligonucleotide substrates for aminoacylation. All substrates terminate at the 3′-end in the single-stranded NCCA tetra nucleotide, where the identity of N depends on the specific aminoacylation system. The four nucleotides (N_4) of the loop of the tetraloop structure are specifically chosen to impart stability to the short (4 bp) helix joined to that loop (37). (*Bottom*) The pseudo-knotted RNA circle described by Felden and Giegé (9).

histidyl-tRNA synthetase. These mutations affected the efficiency of aminoacylation in the ways expected, based on results with the full tRNA and with a conventional minihelix structure. So, the lesson in general seems to be that the RNA structural format in which the determinants are presented does not make much difference, as long as those determinants are accessible.

Aminoacylation systems are imagined to have started with ribozymes (26, 27) that gradually gave birth to ribonucleoproteins (like RNase P; refs. 28–30) and then to the contemporary synthetases (31). These early catalysts may have interacted with just the nucleotides at the end of the RNA substrate, as evidenced by the location of contemporary determinants near the amino acid attachment site. Indeed, the primordial, historical tRNA synthetase is thought to be represented by the domain of the contemporary enzyme that contains the catalytic site. This domain has insertions that enable docking of the 3′-end of the tRNA substrate at the active site. A second domain of synthetases, which is believed to be a later addition, interacts with more distal sites in the tRNA structure, including (in many instances) the anticodon.

The diverse artificial substrates studied so far raise the possibility that aminoacylation of RNAs other than tRNAs may be found in contemporary cellular systems. The large

plant viral RNAs that are aminoacylated with specific amino acids are a prominent class of novel substrates (3, 23, 32). Here the 3′-end of the viral RNA recreates a tRNA-like structure through the formation of a pseudoknot that recapitulates the minihelix domain. Another novel structure is 10Sa RNA (33, 34), which has a complex secondary structure (35) that is charged with alanine (the 10Sa RNA contains the G3:U70 base pair that is essential for aminoacylation with alanine). This RNA is both an amino acid acceptor and an mRNA (36). We suspect that many more natural examples of novel RNAs that can be aminoacylated will be discovered. These may give the best insights yet into the development of early systems of aminoacylation, especially if they can be related to the tRNA structure in some way. The challenge is to find them.

1. Francklyn, C. & Schimmel, P. (1989) *Nature (London)* **337,** 478–481.
2. Frugier, M., Florentz, C. & Giegé, R. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 3990–3994.
3. Rudinger, J., Florentz, C., Dreher, T. & Giegé, R. (1992) *Nucleic Acid Res.* **20,** 1865–1870.
4. Nureki, O., Niimi, T., Muto, Y., Kanno, H., Kohno, T., Muramatsu, T., Kawai, G., Miyazawa, T., Giege, R., Florentz, C. & Yokoyama, S. (1993) in *The Translation Apparatus*, eds. Nierhaus, K. H., Franceschi, F., Subramanian, A. R., Erdmann, V. A. & Wittmann-Liebold, B. (Plenum, New York), pp. 59–66.
5. Frugier, M., Florentz, C. & Giegé, R. (1994) *EMBO J.* **13,** 2218–2226.
6. Hamann, C. S. & Hou, Y.-M. (1995) *Biochemistry* **34,** 6527–6532.
7. Martinis, S. A. & Schimmel, P. (1995) in *tRNA: Structure, Biosynthesis and Function*, eds. Soll, D. & RajBhandary, U. L. (Am. Soc. Microbiol., Washington, DC), pp. 349–370.
8. Saks, M. E. & Sampson, J. R. (1996) *EMBO J.* **15,** 2843–2849.
9. Felden, B. & Giegé, R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 10431–10436.
10. Park, S. J. & Schimmel, P. (1988) *J. Biol. Chem.* **263,** 16527–16530.
11. Cusack, S., Berthet-Colominas, C., Biou, V., Borel, F., Fujinaga, M., Hartlein, M., Krikliviy, I., Nassar, N., Price, S., Tukalo, M. A., *et al.* (1993) in *The Translation Apparatus*, eds. Nierhaus, K. H., Franceschi, F., Subramanian, A. R., Erdmann, V. A. & Wittmann-Liebold, B. (Plenum, New York), pp. 1–12.
12. Cusack, S., Yaremchuk, A. & Tukalo, M. (1996) *EMBO J.* **15,** 2834–2842.
13. Francklyn, C., Shi, J.-P. & Schimmel, P. (1992) *Science* **255,** 1121–1125.
14. Martinis, S. A. & Schimmel, P. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 65–69.
15. Felden, B., Florentz, C., Westhof, E. & Giegé, R. (1993) *Biochimie (Paris)* **75,** 1143–1157.
16. Hou, Y.-M. & Schimmel, P. (1988) *Nature (London)* **333,** 140–145.
17. McClain, W. H. & Foss, K. (1988) *Science* **240,** 793–796.
18. Weiner, A. M. & Maizels, N. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7383–7387.
19. 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.
20. Weiner, A. M. & Maizels, N. (1994) *Curr. Biol.* **4,** 560–563.
21. Schimmel, P., Giegé, R., Moras, D. & Yokoyama, S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8763–8768.
22. Noller, H. F. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, New York), pp. 137–156.
23. Giegé, R., Puglisi, J. D. & Florentz, C. (1993) *Prog. Nucleic Acid Res. Mol. Biol.* **45,** 129–206.
24. Felden, B., Florentz, C., Giegé, R. & Westhof, E. (1994) *J. Mol. Biol.* **235,** 508–531.
25. Crothers, D. M., Seno, T. & Söll, D. G. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 3063–3067.
26. Piccirilli, J. A., McConnell, T. S., Zaug, A. J., Noller, H. F. & Cech, T. R. (1992) *Science* **256,** 1420–1424.
27. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. (1995) *Science* **267,** 643–647.

Commentary: Schimmel and Alexander

Proc. Natl. Acad. Sci. USA 95 (1998)   10353

28. Altman, S., Baer, M. F., Bartkiewicz, M., Gold, H., Guerrier-Takada, C., Kirsebom, L. A., Lumelsky, N. & Peck, K. (1989) *Gene* **82,** 63–64.

29. Brown, J. W. & Pace, N. R. (1991) *Biochimie (Paris)* **73,** 689–697.

30. Altman, S., Kirsebom, L. & Talbot, S. (1993) *FASEB J.* **7,** 7–14.

31. Henderson, B. S. & Schimmel, P. (1997) *Bioorg. Med. Chem.* **5,** 1071–1079.

32. Dreher, T. W., Tsai, C.-H., Florentz, C. & Giegé, R. (1992) *Biochemistry* **31,** 9183–9189.

33. Oh, B. K. & Apirion, D. (1991) *Mol. Gen. Genet.* **229,** 52–56.

34. Ushida, C., Himeno, H., Watanabe, T. & Muto, A. (1994) *Nucleic Acid Res.* **22,** 3392–3396.

35. Felden, B., Himeno, H., Muto, A., McCutcheon, J. P., Atkins, J. F. & Gesteland, R. F. (1997) *RNA* **3,** 89–103.

36. Keiler, K. C., Waller, P. R. H. & Sauer, R. T. (1996) *Science* **271,** 990–993.

37. Shi, J.-P., Martinis, S. A. & Schimmel, P. (1992) *Biochemistry* **31,** 4931–4936.