Minireview

Operational RNA Code for Amino Acids in Relation to Genetic Code in Evolution*

Published, JBC Papers in Press, January 12, 2001, DOI 10.1074/jbc.R000032200

Lluis Ribas de Pouplana and Paul Schimmel‡

From the Department of Molecular Biology and Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037

The genetic code is based on the specific aminoacylation of transfer RNAs (tRNAs) by aminoacyl-tRNA synthetases (aaRSs) (1, 2). This reaction links anticodon triplets in tRNAs with specific amino acids. The specificity of the reaction is governed by tRNA identity elements that are recognized by the aminoacylating enzymes (2). The universal distribution and conservation of tRNAs and aaRS imply that they preceded the origin of the three kingdoms of life, Bacteria, Archaea, and Eucarya (3–5). Significantly, nucleotide determinants other than the anticodon triplets are important for aminoacylation efficiency and specificity (6, 7). It is these nucleotides (making up an operational RNA code) that are now seen as important for maintaining a universal genetic code.

Background

Typically, aminoacylation occurs in two steps.

\[
E + AA + ATP \leftrightarrow E\text{-AA-AMP} + PP_i
\]

\[
E\text{-AA-AMP} + tRNA \leftrightarrow E + AA\text{-tRNA} + AMP
\]

**REACTIONS 1 AND 2**

First, the enzyme (E) condenses its cognate amino acid (AA) with ATP to form a tightly bound aminoacyl adenylate (AA-AMP) with the release of pyrophosphate (PPi). The aminoacyl group is then transferred to the 3’-end of tRNA to give aminoacyl-tRNA (AA-tRNA) and release of AMP. In this way, a specific nucleotide triplet (anticodon) in the tRNA is physically connected (through the tRNA structure) with a particular amino acid.

Transfer RNAs are usually comprised of 76 nucleotides arranged into a cloverleaf structure with four major arms. The acceptor stem is a helix of 7 bp that ends on the 3’-side with the universal tetranucleotide NCCA26, with the amino acid attachment site at A76. The dihydrouridine-, TψC-, and anticodon-stem-loop make up the other three arms (Fig. 1). The four arms are arranged in three dimensions into an L-shaped structure (5, 9), where the acceptor and TψC stems stack together to make up a 12-bp hairpin known as the minihelix (ending in the TψC loop) (10). At right angles, the D- and anticodon stems fuse to give a 10-bp helix with the D-loop at one end and the anticodon loop at the other. Thus, the triplet of the code and its cognate amino acid are in distinct domains at opposite ends of the tRNA structure (Fig. 1).

The minihelix domain terminating in the CCA trinucleotide is found as a regulatory element for replication of specific RNA genomes (11, 12). In the ribosome, the anticodon-containing and the minihelix domain bind to distinct tRNAs (13). This observation raises the possibility that the minihelix and anticodon-containing domains had separate origins. That an ancient minihelix duplicated and gave rise to the anticodon-containing domain and genetic code has also been proposed (14).

In bacteria, there typically is one aaRS for each amino acid. In eukaryotes, distinct nuclear encoded cytoplasmic and mitochondrial enzymes carry out aminoacylations in their respective cellular compartments. Broadly speaking, the enzymes are comprised of two major domains. The historical, most ancient domain contains the catalytic site with determinants for binding the minihelix portion of the tRNA. These catalytic domains are limited to two folds that define two families known as classes I and II (15–19). (With rare exceptions, each class contains enzymes specific for 10 different amino acids.) Most of the structural evolution that gave rise to the two classes of synthetases took place before the first split of the universal tree of life based on analyses of 16S RNA sequences (4, 5, 20). The synthetases also have a second major domain that, in many instances, interacts with the anticodon. The idiosyncratic structures of these domains, even for enzymes within the same class, suggest that the second domain was added later in evolution.

**The Minihelix and an Operational RNA Code for Amino Acids**

An obvious way for an aaRS to relate a specific amino acid to a nucleotide triplet is through direct recognition of the tRNA anticodon. However, the anticodon is not used as the principal determinant for aminoacylation by alanyl-, seryl-, or leucyl-tRNA synthetases (1). For example, bacterial and eukaryote cytoplasmic alanyl-tRNA synthetase throughout evolution rely on a specific G3:U70 base pair in the acceptor stem to define the identity of tRNAAsp (21–25) (Fig. 2). No physical contact is made by the enzyme with the anticodon (26). As a consequence, a minihelix or even smaller helices (e.g. microhelices of 7 bp) that contain a G3:U70 base pair are robust substrates for aminoacylation by bacterial, yeast, and human enzymes (10, 24, 25). Variants of these substrates with natural and non-natural base analogs have been useful for evaluating energetic contributions of the G3:U70 base pair (6, 27–29).

These observations are mirrored by numerous examples of tRNA synthetases that charge microhelices based on the sequences of the acceptor stems of their cognate tRNAs (6, 7, 10, 30–44). Thus, despite synthetase contacts with the anticodon (45), the acceptor stem often contains determinants sufficient for specific aminoacylations. The sequences/structures in RNA oligonucleotides that mimic the acceptor stem and confer specific aminoacylations constitute an operational RNA code for amino acids (46) (also referred to as the “second genetic code” (47, 48)). These determinants typically are comprised of 1–3 bp and the N73 “discriminator” base. The operational RNA code may have predated the genetic code and according to some analyses was the progenitor of the genetic code (6, 48–52).

**Barriers to Cross-domain Aminoacylations and Their Manipulation**

The tyrosine and glycine systems illustrate how the position of acceptor stem determinants for aminoacylation, but not the determinants themselves, have been conserved (1, 2). For example, euabacterial TyrRS do not aminoacylate eukaryotic cytoplasmic tRNA Tyr5 (53). Conversely, euahbacterial tRNA Tyr cannot be aminoacylated by eukaryotic TyrRS. This domain specificity correlates with the change of the conventional G1:C72 base pair found in most tRNAs to C1:G72 in eukaryotic and archael tRNA Tyr5 sequences.2

The 1:72 base pair was demonstrated to be important for aminoacylation of microhelices or tRNAs based on sequences of tyrosine acceptors in Bacillus stearothermophilus (55), the eukaryote pathogen Pneumocystis carinii (56), the yeast Saccharomyces cerevisiae

‡ To whom correspondence should be addressed: EJC-379, Scripps Research Inst., 10550 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-784-0890; Fax: 858-784-0900; E-mail: schimmel@scripps.edu.

The abbreviations used are: aaRS, aminoacyl-tRNA synthetase(s); bp, base pair(s).

---

* This minireview will be reprinted in the 2001 Minireview Compendium, which will be available in December, 2001. This work was supported by Grant 23562 from the National Institutes of Health and by a fellowship from the National Foundation for Cancer Research.

† The National Foundation for Cancer Research.

‡ To whom correspondence should be addressed: EJC-379, Scripps Research Inst., 10550 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-784-0890; Fax: 858-784-0900; E-mail: schimmel@scripps.edu.

‡ To whom correspondence should be addressed: BCC-379, Scripps Research Inst., 10550 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-784-0890; Fax: 858-784-0900; E-mail: schimmel@scripps.edu.

* This minireview will be reprinted in the 2001 Minireview Compendium, which will be available in December, 2001. This work was supported by Grant 23562 from the National Institutes of Health and by a fellowship from the National Foundation for Cancer Research.

† To whom correspondence should be addressed: EJC-379, Scripps Research Inst., 10550 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-784-0890; Fax: 858-784-0900; E-mail: schimmel@scripps.edu.

The abbreviations used are: aaRS, aminoacyl-tRNA synthetase(s); bp, base pair(s).
Minireview: Operational RNA Code for Amino Acids

A Variation That Suggests Relative Timing of Appearance of Synthetases and tRNAs

Unlike most species, the archaeabacterium Methanococcus jannaschii does not have a gene coding for a class II lysyl-tRNA synthetase (60). Instead, aminoacylation of tRNA\(^{\text{Asn}}\) is catalyzed by a class I enzyme (61). Phylogenetic analysis of the novel class I LysRS showed that its origin cannot be explained by a recent gene transfer event (62). Analysis of sequences of tRNA\(^{\text{Asn}}\) from all phylogenetic domains showed that tRNA\(^{\text{Asn}}\) does not divide into two groups that follow the distribution of its two different aminoacylating enzymes (62). The coherence of the tRNA\(^{\text{Asn}}\) sequences implies that the identity of this tRNA was established independently (and probably before) the establishment of the two forms of LysRS (62). This situation is unlike the case of glutaminyl- and asparaginyl-tRNA synthetases. These two aaRS appeared later in evolution as result of duplications of genes coding for AspRS (63). We propose that the lack of documented examples of such transfers may be related to the ability of AspRS to recognize the related tRNA\(^{\text{Asp}}\) in certain organisms. For example, in archaea, a canonical asparagine-tRNA synthetase is missing. The aminoacylation of tRNA\(^{\text{Asp}}\) with asparagine is accomplished through an initial asparylation of tRNA\(^{\text{Asp}}\) catalyzed by AspRS. This asparylation is followed by a transamidase catalyzed by a separate enzyme (72).

Not surprisingly, archaeal tRNA\(^{\text{Asn}}\) contains the important G73. To recognize tRNA\(^{\text{Asp}}\), archaeal AspRS has a modified recognition mechanism for the anticodon. In particular, it is insensitive to the base at position 36 (the only anticodon difference between tRNA\(^{\text{Asn}}\) and tRNA\(^{\text{Asp}}\)) (73).

Bacterial and eukaryotic organisms do not require this mechanism of generating Asn-tRNA\(^{\text{Asn}}\), because they utilize a canonical AsnRS that probably arose as a duplication of an ancient AspRS. Thus, the transamidase is found in bacteria and archaea. The transamidase catalyzes the formation of Asn-tRNA\(^{\text{Asp}}\) via a transamidase reaction (71). Indeed, the transamidase is found in Thermus thermophilus (71).

Subtle Variations in Operational RNA Code May Be Essential to Maintain a Universal Genetic Code

At the time of emergence of the translational apparatus, the operational RNA code had the capacity to adapt to the problems of discrimination of increasingly large populations of RNA molecules. These populations extended beyond just tRNAs to cellular RNAs such as mRNAs that could potentially cross-bind and thereby inhibit a tRNA synthetase. Identity elements in tRNA acceptor stems mutated in different taxonomic groups, preventing cross-species aminoacylations of many different tRNAs. These aminoacylation barriers blocked genetic exchanges involving genes for tRNAs and their synthetases and were probably important to avoid disruption of the genetic code. In particular, if two distinct synthetases for the same amino acid are present (either via gene duplication or because of lateral gene transfer), then mutations can accumulate in one of them while the other is held fixed. It is understood that gross mischargeing would be rapidly eliminated, but more subtle interactions with suppressor tRNAs or tRNAs containing infrequent codons could alter the amino acid or tRNA specificity and gradually introduce changes in the codon-amino acid relationships (75).

Disruption of the universal genetic code has happened in rare instances, but its overall conservation points to the existence of strong selection pressures against its variation. In this context, the operational RNA code for aminoacylation of tRNA molecules would
act as a strong deterrent against contamination of the code through nonspecific charging. This requirement for tight tRNA recognition might explain why most tRNA synthetases in bacteria and archaea are encoded by single copy genes. With single copy genes, the opportunity for contamination of the genetic code is greatly restricted. The situation in eukaryotes is essentially the same. A separate, distinct gene is designated for a synthetase in each cell compartment, the cytoplasm and mitochondria.

**Operational RNA Code in Relation to Emergence of Eukaryotic Cell**

The combination of archaeabacterial and eubacterial species gave rise to the eukaryotic cell and generated organelles like plastids and mitochondria (76). The physiologic fusion of these species had to include integration of their systems for aminoacylation. The evolutionary solution to the duplicated synthetase-tRNA systems could have been determined by the interplay between mutations in the synthetases (and acceptor stem elements) that were being merged (54, 64). Eventually, two compartments, mitochondria and cytoplasm, emerged that utilized the same genetic code.

For example, two genes for the same enzyme activity (cytoplasmic and mitochondrial) are encoded by the genomes of contemporary eukaryotes. The possibility for two versions of the same enzyme ending up in the same cellular compartment is consequently significant. Were a misplaced mitochondrial enzyme to recognize the same acceptor stem elements as its cytoplasmic counterpart, the presence of two synthetases in the cytoplasm (targeted to the same acceptor stem) gives opportunity to invade and alter the genetic code (see above). By having distinct recognition elements for the mitochondrial form, the likelihood of such an invasion is greatly diminished. This consideration may account in part for why acceptor stem elements for animal mitochondria show more differences from their cytoplasmic counterparts than is typically seen for the same elements compared across taxonomic domains. Thus, although throughout evolution the G3:U70 base pair marks a tRNA for aminoacylation with alanine (Fig. 2), G3:U70 is often not found...
In animal mitochondrial tRNA\textsuperscript{Ala}. (However, G3:U70 is commonly found in mitochondria of other eukaryotes.)\textsuperscript{2} Analysis of identity elements for other animal mitochondrial tRNA sequences revealed that tRNA\textsuperscript{Ala} is not an isolated example.

The striking variations in animal mitochondrial tRNA identity elements may also be related in part to a decreased level of genome complexity. In general, the set of tRNA genes in animal mitochondria is largely reduced (E. coli contains 40 tRNA genes and the human mitochondria contains 22). This reduction means that tRNA\textsuperscript{Ala} synthetases now have to discriminate among a smaller population of tRNA molecules. Perhaps, under these circumstances, the evolutionary pressure to maintain a given set of identity elements is reduced, because certain discrimination problems no longer exist.

In summary, the genetic code is seen as preserved throughout evolution (across all taxonomic domains and in higher eukaryotes with their separate cell compartments) as a consequence of adaptations in the identity elements in tRNA acceptor stems that constitute an operational RNA code. These adaptations are necessary to the need to keep anticodon sequences fixed to a universal code, on the one hand, and on the other, the need to facilitate the expansion and diversification of living organisms. This ancient RNA code, which may have started with the sequence-specific aminoacylation of minihelix-like precursors of tRNAs by ribozymes, has endured long after the genetic code was established because it offered a defense against invasions of the code arising from tRNA synthetases with amino acid-anticondor assignments that differed from those of the universal code. It also endured because of its capacity to respond to increasingly large and diverse populations of RNAs and the problems of discrimination that they presented.

REFERENCES

1. Giegé, R., Sissler, M., and Florentz, C. (1998) Nucleic Acids Res. 26, 5017–5035
2. Beach, P. J., and Musier-Forsyth, K. (1999) Biochemistry 38, 1–28
3. Nagel, G. M., and Doolittle, R. F. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 8121–8125
4. Nagel, G. M., and Doolittle, R. F. (1995) J. Mol. Biol. 40, 487–498
5. Brown, J. R., and Doolittle, W. F. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 2441–2445
6. Musier-Forsyth, K., and Schimmel, P. (1999) Acc. Chem. Res. 32, 368–375
7. Martinis, S. A., and Schimmel, P. (1996) in Escherichia coli and Salmonella Translation Apparatus (Nierhaus, K. H., Franceschi, F., Subramaniam, A. R., Erdmann, V. A., and Wittmann-Liebold, B., eds) Plenum Press, New York
8. Martinis, S. A., and Schimmel, P. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 8763–8767
9. De Duve, C. (1991) Blueprint for a Cell: The Nature and Origin of Life, Neil Patterson, Burlington, NC
10. Woese, C. R. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 337–343
11. Woese, C. R., Olsen, G. J., Ibba, M., and Soll, D. (2000) Science 290, 137–156
12. Weiner, A. M., and Maizels, N. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 2095–2104
13. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1997) Science 278, 278–281
14. Saks, M. E., and Samson, J. R. (1992) EMBO J. 11, 2219–2226
15. Poole, R. P., and Schimmel, P. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 12124–12129
16. Felden, B., and Giegé, R. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 10431–10436
17. Alexander, R. W., Nordin, B. E., and Schimmel, P. (1997) EMBO J. 16, 1–10
18. Alexander, R. W., Nordin, B. E., and Schimmel, P. (1998) J. Mol. Biol. 281, 249–258
19. Cusack, S. (1995) Trends Biochem. Sci. 20, 262–266
20. Moras, D. (1992) J. Biol. Chem. 267, 21455–21462
21. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1997) Science 278, 278–281
22. Poole, R. P., and Schimmel, P. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 12124–12129
23. Hou, Y. M., and Schimmel, P. (1999) Science 286, 6069–6072
24. Nureki, O., Niimi, T., Muto, Y., Kanno, H., Kohno, T., Muramatsu, T., Kawai, G., Miyazawa, T., Giegé, R., Florentz, C., and Yokoyama, S. (1993) in The Translation Apparatus (Nierhaus, K. H., Franceschi, F., Subramaniam, A. R., Erdmann, V. A., and Wittmann-Liebold, B., eds) Plenum Press, New York
25. Sampson, J. R., and Saks, M. E. (1993) Nucleic Acids Res. 21, 4467–4475
26. Prugger, M., Florentz, C., and Giegé, R. (1994) EMBO J. 13, 2219–2226
27. Prugger, M., Florentz, C., and Giegé, R. (1994) J. Mol. Biol. 236, 710–724
28. Musier-Forsyth, K., and Schimmel, P. (1992) Proc. Natl. Acad. Sci. U. S. A. 90, 3470–3474
29. Musier-Forsyth, K., Shi, J.-P., Henderson, B., Bald, R., Jöf, J. P., Erdmann, V. A., and Schimmel, P. (1995) J. Am. Chem. Soc. 117, 7523–7524
30. Frangklyn, C., and Schimmel, P. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 8655–8659
31. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125
32. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125
33. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125
34. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125
35. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125
36. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125
37. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125
38. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125
39. Franklyn, C., Shi, J., and Schimmel, P. (1992) Science 255, 1121–1125