SPECULATIONS ON THE EVOLUTION OF THE GENETIC CODE III: THE EVOLUTION OF t-RNA

Hyman Hartman

Department of Earth, Atmospheric, and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

It is postulated that the primitive genetic code was a doublet G, C code. The code consisted of GG coding for gylcine, CC coding for proline, CG coding for arginine and GC coding for alanine. The major speculation of this paper is that the remnants of this primitive code exist in the loops of the present day t-RNA. If one considers the cloverleaf structure of t-RNA, then aside from the anti-codon loop, the major invariances of the loops or "open loops" are $5'UGGU_3'$, $5'T\psi CGA_3'$, $5'*CCA_3'$. The GG, CG, CC are considered to be remnants of the ancient doublet G,C code. Furthermore, modern t-RNA is a "tetramer" of a "monomer" (arm and loop structure) and it is in the loop that the evolution of t-RNA can be followed. Balasubramanian suggested that the anti-codon loop structure $5'U...A_3'$ was the earliest t-RNA. In this paper we speculate that the pentamer structure was preceded by UGCA, UGGA, UCCA and UCGA. It is further postulated that these ancient t-RNAs were preceded by UGC, UGG, UCC and UCG. Thus the modern t-RNA has had a long history and part of it can be read in its present structure. A more speculative assumption is that the 5' uracil was a thio-uracil. The activated amino acids were in the thio-ester form and later evolved into the ester form found on the 2, or 3 hydroxyl of the adenosine ribose.

On confronting a complex biological system, such as the protein synthesizing machinery, it is a useful strategy to speculate on how it may have evolved from simpler precursors. This has been done with respect to the evolution of the genetic code (6) and metabolism (7). The previous two papers on the evolution of the genetic code postulated a primitive GC doublet code in which GG coded for glycine, and CC, GC, CG coded for proline,

Origins of Life **14** (1984) 643–648. 0302-1688/84/0141-0643$00.90

alanine and arginine respectively. This paper is the first in a series which will deal with the origin and evolution of the protein synthesizing machinery.

The translation apparatus involves messenger RNAs, transfer RNAs, (t-RNA) ribosomes and amino acid synthetases. The many RNAs, messenger, ribosomal and t-RNAs involved, must have an evolutionary explanation. The t-RNAs because of their small size would appear to be the most primitive from the evolutionary point of view. Grosjean et al. (3) have studied the t-RNA and t-RNA interaction when the anti-codons were complimentary. Their main conclusions reached were that the seven nucleotides in the anti-codon loop constrained and enhanced the binding of the t-RNAs and that the stacking of the anti-codon nucleotides on the two t-RNAs led to enhanced binding. They then speculated that "the interaction between two RNA loops may have been part of the physical basis for the evolutionary origin of the code and that this mechanism may still be utilized by folding the m-RNA on the ribosome into a loop similar to the anti-codon loop" (4). The inferred symmetry between the anti-codon loop and the messenger RNA-loop suggests that the messenger RNA may have the ability to accept amino acids as do t-RNAs. This is precisely what has been found in some plant viruses (5). These viral RNAs function as messenger RNA. At the 3' end of the virus RNA there ixist a CCA sequence which accepts an amino acid (in the presence of the amino acid t-RNA synthetases); an example is Tobacco Mosaic Virus (TMV) which accepts a histidine. The amino acid is in an ester linkage with either the 2' or 3' hydroxyl group of the ribose of adenosine. In the comparison between the t-RNA of histidine and the 3' terminal sequence of TMV the loops postulated have similar structures especially that loop found between 25-35 which has the sequence UAAGCUU which similar to xx AGC ψ T found in the histidine t-RNA (Hall 1979). The conservation of the hairpin loops suggests a common origin for both the t-RNA and the TMV RNA.

In the replication of small single-stranded RNA bacteriophages (MS 2 and Q β), the RNA replicase contains four components; 1) protein S1 found in the 30S ribosome, 2) EF-Tu, 3) EF-Ts elongation factors involved in protein synthesis and, 4) a viral encoded polypeptide (2). Since t-RNAs are recognized by the elongation factors, it has been suggested that the t-RNA like structures found at the 3' end of the viral RNAs are involved in the initiation of RNA replication.

The simplest hypothesis which explains the role which t-RNA structures play in the replication of viral RNA and the amino acid acylation of viral messenger RNAs, is that t-RNA and messenger RNA are evolved from a common ancestor. Woese first proposed that a precursor to both the ribosomal RNA and transfer RNA was a small oligonucleotide subunit which had a loop and a double helical

strand 5 bases long (12). Recently Mizutani and Ponnamperuma (9) have suggested that "the ancestral form of t-RNA could have been composed of seventeen nucleotides of which seven nucleotides formed a loop and of which two pentanucleotides at either end of the loop constituted complementary double helical structure". We will assume that such a structure was a precursor not only to the transfer RNA but to the ribosomal RNA and to messenger RNA as well.

Replication would have been a hierarchial process. First the synthesis of many different ants (Ancestral t-RNA). The ants would have been joined together by a primitive RNA ligase. The catalyst RNA ligase splices two RNA molecules by forming 3'-5' phosphodiester bond between the 3' hydroxyl of one RNA molecule and the 5' phosphate of another RNA molecule (11). It is therefore relatively easy to use a polymerized set of primitive t-RNAs as a template for the polymerization of a complementary set of primitive t-RNAs (ants) which would be polymerized by a primitive RNA ligase. The same complimentarity between the triplets or doublets in the loops between the ants would be used in translation and it would be no accident that replication and translation were similar processes. It should be pointed out that when these ants interact through their loop structures there is a two-fold rotational symmetry axis between the middle bases. This means that the coding and the anticoding system were identical unlike the system today where the coding system messenger RNA is quite distinct from the anticoding system t-RNA and ribosomes. At some point in evolution, the symmetry between the coding system and the anticoding system was broken.

The cloverleaf structure of t-RNA can thus be considered as four ants jointed together. The complex tertiary structure of the present t-RNA records the evolution of the protein synthesizing machinery. The four components of the t-RNA cloverleaf are 1) the anticodon loop and arm, 2) the TψCGA loop and arm, 3) the UGGU loop and arm, and 4) the XCCA end (this is a loop which has been opened) and the amino acid accepting arm. The loop and arm structure suggests that the loop itself was the most ancient t-RNA. This observation and others have led Balasubramanian to postulate that "oligonucleotides of five residues having U at the 5' end, a purine at the 3' end and any combination of three bases in the middle is taken as a primitive t-RNA" (1). If we examine the TψCGA structure in the TψGCA loop, it was pointed out by Woese (13) that this structure could pair with itself as follows:

5' UCGA 3'
3' AGCU 5'

U is substituted for ψ, as it is considered as the evolutionary precursor. Thus it is possible to postulate a simpler

oligonucleotide as the primitive t-RNA, an oligonucleotide of four residues having U at the 5' end and adenosine at the 3' end and any combination of G and C, taken two at a time in between.

In the glycine t-RNA of prokaryotes the four bases at the end of the amino acid accepting arm are UCCA. It is thus inferred that a set of primitive t-RNA with codons and anticodons of the form UGGA, UCCA, UGCA, UCGA preceded the more complex triplet code. The adenosine in the case of UCCA and the others was the site where the amino acids, glycine, proline alanine and arginine formed an ester with either the 2' or 3' hydroxyl of ribose. The polypeptides which would be coded for are best exemplified today by protamines. The ancestral peptides of the protamines are believed to be duplications of a pentapeptide, composed of one alanine and four arginines (AR_4). It has been speculated that "the ancestral pentapeptide AR_4 and its immediate successor AR_4 AR_4 probably date from an era before the existence of DNA as we know it. For instance, this peptide could be coded by an ancestral "DNA", consisting of GC pairs (10). The remnants of this GC code can still be seen in the CC, GG, and GC doublets to be found in the amino acid accepting oligonucleotide UCCA, the GG in the UGGU loop and the GC in the TΨCGA loop, all found at the ends of the three arms of the present day t-RNAs.

The UGG found in UGGU structure implies that the A at the 3- end of the t-RNA was a later addition. Thus one might simplify the codon and anticodons further (i.e., UGG, UCC, UCG, UGC). The uridine which precedes the doublet may have been a thiouridine which could have formed a thioester with the amino acid. It has been found that in the peptide antibiotics, the thioesters of the amino acids were the activated intermediates in the polymerization (8). It is thus speculated that the first activated amino acids involved in protein synthesis were thioesters rather than adenylated amino acids.

To summarize, the code began as a doublet GC code, the codons and anticodons were CC, GG, CG, GC which were preceded by a thiouridine. The activated amino acids were thioesters associated with the thiouridine. This system was succeeded by one in which an adenosine was added to the 3' end giving UGGA, UCCA, UGCA, UCGA. The replication and translation apparatus had not separated and the activated amino acids were now esters rather than thioesters. The peptides were small about 4 or 5 amino acids in length. The codons and anticodons became five nucleotides in length U...A and the triplets rather than doublets now coded for amino acids. Finally, the loop and arm structure evolved and polymerized into the form of t-RNA as we know it today.

The system in which the t-RNA evolved is postulated to be a replicating clay system. In a previous paper (6) the lipid

bilayer formed between clay layers, is considered the site where the genetic code originated. The next paper will deal with this system.

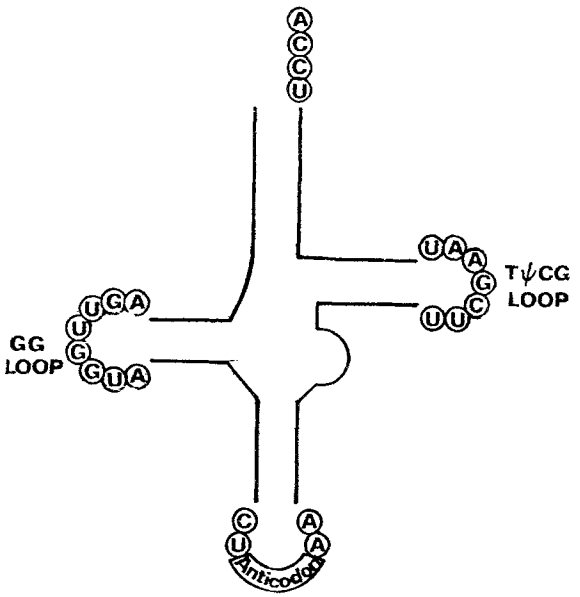

Figure 1

The nucleotide sequences in the loops are adapted from Eigen's reconstruction of the ancestral t-RNA.

References

1. Balasubramanian, R., Seetharamulu, P. and Raghunathan, G., 1980. Origins of Life, 10, 15.

2. Blumenthal, T. and Carmichael G.G., 1979. Ann. Rev. Biochem, 48, 525.

3. Grosjean, H., Soll, D.G. and Crowthers, D.M., 1976. J. Mol. Biol. 103, 499.

4. Grosjean, H., De Henau, S. and Crowthers, D.M., 1978. PNAS 75, 610.

5. Hall, T.C., 1978. Int. Rev. of Cytology, 60, 1.

6. Hartman, H., 1975a. Origins of Life, 6, 423.

7. Hartman, H., 1975b. J. Mol. Evol. 4, 359.

8. Lipmann, F., 1971. Science, 173, 875.

9. Mizutani, H. and Ponnamperuma, C., 1980. Origins of Life, 10, 31.

10. Phillips, D.M.P., 1971. In: Histones and Nucleohistones, Ed. D.M.P. Phillips, Plenum Press, New York, p. 69-70.

11. Uhlenbeck, O.C. and Gumport, R.I., 1982. In: The Enzymes, ed. by P.D. Boyer, Academic Press, New York, p. 31.

12. Woese, C., 1969. Progress in Molecular and Subcellular Biology, Vol. 1, 5.

13. Woese, C., 1970. Nature, 226, 817