ACCOUNTS OF CHEMICAL RESEARCH

VOLUME 10 NUMBER 11 NOVEMBER, 1977

Special Issue on Transfer RNA

Introduction to Transfer RNA

Alexander Rich* and Paul R. Schimmel*

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

Received July 6, 1977

The transfer RNA (tRNA) molecule is an ancient component of biological systems. It probably arose some 4 billion years ago in the early prebiotic period, during which time the increasing development of complex organic molecules led to the creation of polynucleotide chains which had the capacity to contain information as well as to carry out molecular self-replication.

Although the nucleic acids store genetic information, they are unable to express it. The major mode for expressing genetic information is the synthesis of proteins. Proteins provide a large variety of complex chemical environments and have a diversity of biochemical functions that are the molecular basis of living systems.

The tRNA molecule acts at the crossroads between the information-containing polynucleotide chains and the proteins which express genetic information. At one end of the tRNA molecule the three anticodon bases interact with messenger RNA; at the other end the molecule is attached to the growing polypeptide chain during protein synthesis. This molecule is possibly as ancient a component of biological systems as the system of expressing genetic information through the polymerization of amino acids.

Twenty different amino acids are found in contemporary biological systems, and they are associated with 20 different enzymes (aminoacyl-tRNA synthetases) which aminoacylate tRNA molecules.[1,2] Each aminoacyl-tRNA synthetase combines with one of a specific group of tRNA molecules and catalyzes the formation of an ester bond between a particular amino acid and the ribose at the 3′-hydroxyl end of the tRNA molecule. In general, more than one species of tRNA can act as an acceptor for each amino acid in this enzymatic process. These are called the isoacceptor families; however, there appears to be only one aminoacylating enzyme for each amino acid.

The essential feature of this system is the high specificity of aminoacylation. For example, the leucine aminoacylating enzyme has the capacity to aminoacylate all the leucine tRNAs (of which there are about five in *E. coli*) with a high degree of fidelity and rarely if ever makes the mistake of aminoacylating any of the other isoacceptor families. In order for this process to take place, the aminoacylating synthetases must be able to differentiate one species of tRNA from another with high fidelity. Thus, the tRNAs must all have important elements of uniqueness.

After leaving the surface of the synthetase, the tRNA with its attached amino acid makes its way into the ribosomal structure in which protein synthesis takes place. The aminoacyl-tRNA is bound to an elongation factor which enables the charged transfer RNA to interact with the ribosomal and messenger RNA machinery. In the ribosome the three anticodon bases of the tRNA molecule interact with the complementary codon triplet of bases of messenger RNA in the aminoacyl (A) site of the ribosome. This site is immediately adjacent to another tRNA molecule with an attached peptide chain which is already in the ribosome at the peptidyl (P) site.

The mechanics of what goes on in the ribosome are not clear, but the net effect of the molecular mechanism is such that a ribosomal enzyme, the peptidyl transferase, cleaves the growing polypeptide chain from the peptidyl-tRNA and transfers it to the α-amino group

(1) L. L. Kisselev and O. O. Favorova, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 40, 141 (1974).
(2) D. Söll and P. R. Schimmel, *Enzymes, 3rd Ed.*, 10, 489 (1974).

of the aminoacyl tRNA, thereby elongating the polypeptide chain by one amino acid added to its carboxyl terminus. Following this event, both the newly formed peptidyl-tRNA and the messenger RNA are translocated from the aminoacyl (A) site into the peptidyl (P) site, and the system is ready again for another cycle in protein synthesis. There appear to be no constraints on the sequence of amino acids in protein synthesis, and it is apparent that virtually all tRNAs are designed to go through this ribosomal machinery. Thus, the various tRNAs must have important elements of commonality.

In a sense, there is a paradox in trying to understand this central biochemical function of tRNA. On the one hand, the tRNA molecules must be sufficiently different to be discriminated by different aminoacyl synthetases. At the same time they must be sufficiently alike so that they can all go through the same ribosomal apparatus during protein synthesis. We would like to understand the manner in which this twofold chore is carried out in an effective and relatively error-free mode.

In 1965 Holley and co-workers reported the sequence of the first tRNA molecule.[3] They noted that there are segments of the polynucleotide chain which appear complementary if the chain is folded back upon itself. One of these foldings has given rise to the cloverleaf diagram (Figure 1) in which sequences are represented as a series of stems and loops, the stems being composed of complementary bases with Watson–Crick hydrogen bonding similar to those which are found in the familiar DNA double helix. As additional sequences were reported, it became apparent that this cloverleaf folding is expressing something of a fundamental nature concerning the molecule.

Transfer RNA molecules contain about 73 to 93 nucleotides in a single polynucleotide chain. Up to the present, over 80 different tRNA molecules have been sequenced, and Figure 1 summarizes the available information[4-6] for all sequences except initiator tRNAs. Nucleotides are indicated as open circles, except where certain characteristic nucleotides are found in certain positions. Hydrogen bonds between bases are indicated by straight lines in the stem regions of the diagram.

There are a number of constant features. The acceptor stem generally has seven base pairs, and there are four residues at the 3′ end of the chain which terminates in the constant sequence C-C-A.[7] The 3′-terminal adenosine contains the ribose to which the amino acid is attached. The T-ψ-C[7] stem contains five base pairs, and there are seven nucleotides in the loop. This is the most highly specified stem and loop. Those nucleotides which are constant in particular positions for chain-elongating tRNAs are indicated by letters in Figure 1. The preservation of certain nucleotide sequences (e.g., G-T-ψ-C, etc.) in a large number of different molecules suggests that they may be there for reasons related to the three-dimensional conformation of the molecule or because they provide a constant mode of interaction with other molecules.

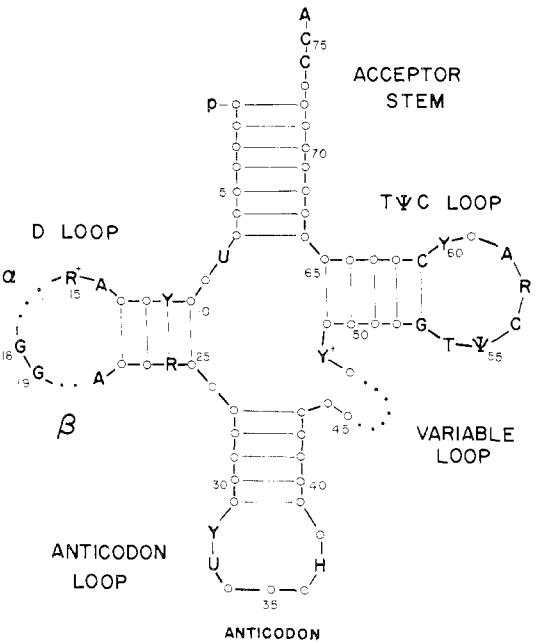

Figure 1. A diagram of all tRNA sequences except for initiator tRNAs. The position of invariant and semiinvariant bases is shown. The numbering system is that of yeast $tRNA^{Phe}$. Y stands for pyrimidine, R for purine, H for a hypermodified purine. R_{15} and Y_{48} are complementary. The dotted regions α and β in the D loop and the variable loop contain different numbers of nucleotides in various tRNA sequences.

In three regions in the polynucleotide chain variable numbers of nucleotides are found (see Figure 1). These are the variable loop and the α and β regions in the dihydrouracil (D) loop. The variable loop contains four or five nucleotides in approximately 80% of the tRNA sequences, but 13–21 nucleotides in the remaining 20%. The α and β regions in the D loop contain from 1 to 3 nucleotides in different tRNA species. They are usually pyrimidines, and many of the pyrimidines have been modified to dihydrouracil residues. Within an isoacceptor family, there are often some species which have different sequences of residues in the α and β regions. The biological roles of the variable number of residues in the variable loop as well as of the α and β regions are unknown at the present time.

The anticodon stem contains five base pairs, and there are seven nucleotides in the loop. The central three nucleotides are the anticodon. The D stem contains three or four complementary base pairs, and the loop has variable numbers of residues depending upon the size of the α and β segments. There are several constant features in the D loop, including the two invariant guanine residues. Figure 1 shows that altogether 23 positions in tRNAs are either invariant (constant bases) or semiinvariant (either purine or pyrimidine). The reasons for this large number of conserved residues remained unknown until the elucidation of the three-dimensional structure of yeast phenylalanine tRNA.[8]

Besides their roles in protein synthesis, tRNAs also participate in many other cellular processes. For many years there have been data suggesting that tRNAs are involved in the regulation of gene expression, partic-

(3) R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisse, S. H. Merrill, J. R. Penwick, and A. Zamir, *Science*, **147**, 1462 (1965).
(4) H. G. Zachau, *Angew. Chem., Int. Ed. Engl.*, **8**, 711 (1969).
(5) B. G. Barrell and B. F. C. Clark, "Handbook of Nucleic Acid Sequences", Joynson-Bruvvers, Ltd., Oxford, 1974.
(6) A. Rich and U. L. RajBhandary, *Annu. Rev. Biochem.*, **45**, 805 (1976).
(7) The notation for the nucleotide bases is as follows: A, adenosine; T, thymidine; C, cytidine; U, uridine; G, guanosine; ψ, pseudouridine; D, dihydrouridine; and Y, highly modified purine. Also, m denotes methyl.
(8) S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H.-J. Wang, N. C. Seeman, and A. Rich, *Science*, **185**, 435 (1974); J. D. Robertus, J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B. F. C. Clark, and A. Klug, *Nature (London)*, **250**, 546 (1974).

ularly of those genes that specify enzymes for the biosynthesis of specific amino acids.[9-11] One of the best-documented cases is the histidine operon studied by Ames and co-workers. In this case the change of two uridines to pseudouridines in the anticodon stem and loop of $tRNA^{His}$ is necessary in order to bring about the regulatory action in the histidine operon.[12] Regulatory roles for tRNA in the biosynthesis of other amino acids are also found.[9-11]

Another example of the diverse activity of transfer RNA is found in polynucleotide synthesis. The reverse transcriptase from two avian tumor viruses uses $tRNA^{Trp}$ as a primer for DNA synthesis off of the viral RNA template.[13,14] It has been demonstrated that reverse transcriptase specifically recognizes $tRNA^{Trp-15}$. Thus, a specific tRNA plays a critical role in the propagation of these tumor viruses. As an additional example, it is known that high molecular weight viral RNAs can be aminoacylated at their 3′ ends with specific amino acids.[16-18] This fact and other data suggest that the 3′ ends of some viral RNAs have tRNA-like structures, and it is plausible to assume that this structure assumes some important biological role.

The transfer RNAs also participate in aminoacyl transferase functions.[19] These reactions include the addition of an amino acid to the N-terminus of a preformed protein[20] or amino acid transfer to acceptors that ultimately are components of cell membranes and cell walls.[21,22]

It is clear from the foregoing that tRNAs are involved in a wide diversity of cellular functions. This reflects the efficient use by nature of a molecule that arose early in evolution. It is likely that many more roles for tRNAs will be found in the future. For example, considerable attention has been given to the idea that tRNAs may play a key role in cellular differentiation, development, and cancer.[10,23,24] It is clear that major research efforts in the future will be directed at exploring and understanding in depth the many roles fulfilled by tRNA.

Even though we are now aware that tRNAs participate in a diversity of cellular processes, the molecule itself presents us with many mysteries. For example, tRNA contains a number of modified nucleosides such as 2′-*O*-methyl-G, ψ, and h_2U (D), N^6-isopentenyladenosine, and 7-methyl-G.[7] Over 50 modifications have been identified. In most cases the biological roles of these modified bases are not known. It is possible that some of these act as signals for specific proteins or other macromolecules that are involved with regulation processes.

Two other questions concern the role of the variable loop and of the variable α and β regions in the dihydrouridine loop. As mentioned above, in approximately 20% of known tRNA sequences the variable loop is large (13 to 21 nucleotides). Since the large variable loop apparently has no role in stabilizing the three-dimensional conformation, the question may be raised as to whether species with large variable loops participate in cellular processes which are not accessible to the simpler tRNAs with only four or five nucleotides in this loop. It is possible that large variable loops serve as receptors for specific proteins. Similar speculations arise concerning the α and β regions in the D loops. These intriguing and somewhat mysterious structural features, together with the known functions and interactions of tRNAs, present challenging areas for future research.

(9) J. M. Calvo and G. R. Fink, *Annu. Rev. Biochem.*, **40**, 943 (1971).
(10) U. Z. Littauer and H. Inouye, *Annu. Rev. Biochem.*, **47**, 439 (1973).
(11) J. E. Brenchley and L. S. Williams, *Annu. Rev. Microbiol.*, **29**, 251 (1975).
(12) C. E. Singer, G. R. Smith, R. Cortese, and B. N. Ames, *Nature (London), New Biol.*, **238**, 72 (1972).
(13) F. Harada, R. C. Sawyer, and J. E. Dahlberg, *J. Biol. Chem.*, **250**, 3487 (1975).
(14) L. C. Waters, B. C. Mullin, T. Ho, and W. K. Yang, *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 2155 (1975).
(15) A. Panet, W. A. Hazeltine, D. Baltimore, G. Peters, F. Harada, and J. E. Dahlberg, *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 2535 (1975).
(16) P. Yot, M. Pinck, A.-L. Haenni, H. M. Duranton, and F. Chapeville, *Proc. Natl. Acad. Sci. U.S.A.*, **67**, 1345 (1970).
(17) S. Litvak, A. Tarragó, L. Tarragó-Litvak, and J. E. Allende, *Nature (London), New Biol.*, **241**, 88 (1973).
(18) R. Salomon and U. Z. Littauer, *Nature (London)*, **249**, 32 (1974).
(19) R. L. Soffer, *Adv. Enzymol.*, **40**, 91 (1974).
(20) R. L. Soffer, *Mol. Cell. Biochem.*, **2**, 3 (1973).
(21) R. M. Gould and W. J. Lennarz, *J. Bacteriol.*, **104**, 1135 (1970).
(22) J. L. Strominger, *Harvey Lect.*, **64**, 179 (1970).
(23) N. Sueoka and T. Kano-Sueoka, *Prog. Nucleic Acid Res. Mol. Biol.*, **10**, 23 (1970).
(24) *Cancer Res.*, **33** (1973).