Application of the Common Features of Transfer RNAs to the Determination of their Nucleotide Sequences

Arthur M. Lesk
Department of Biochemical Sciences
Princeton University, Princeton, New Jersey

Received August 23, 1971

Abstract

The sequences of the fragments produced by T1 ribonuclease digestion of $tRNA_3^{glu}$ of Yeast have been reported. We have determined, with a small amount of remaining ambiguity, an assembly of these fragments that fits the observed regularities in tRNA sequences.

Determinations of the nucleotide sequences of a number of transfer RNA molecules have revealed features common to most or to all tRNAs.[1] All tRNAs possess regions of complementary bases consistent with the cloverleaf secondary structure. Certain positions contain the same base in every tRNA; other positions contain only purines or only pyrimidines. The sequences CCA-3'OH and GTψC-purine occur at the same positions in every tRNA. Modified or "odd" bases also tend to recur at or near regular positions. The purine at position 17 and the pyrimidine at position 63 are complementary in all tRNAs for which sequences have been reported.*

Known tRNA sequences have been determined from the sequences of fragments produced by digestion with T1 ribonuclease (which produces oligonucleotides terminating in 3'G) and pancreatic ribonuclease (which produces oligonucleotides terminating in a 3'-pyrimidine).

We have wondered whether it might be possible to apply the known regularities in tRNA sequences, to deduce the sequence of a tRNA, from a smaller set of data. Yoshida, Takeishi and Ukita have reported the sequences of fragments produces by T1 digestion of $tRNA_3^{glu}$ of Yeast (See Table 1).[2] Is it possible to assemble

* The numbering system is that used by Staehelin.[1] In this numbering scheme, positions 14, 15 and 20 are absent in almost every tRNA -- this must be taken into account in examining the figures in this article.

these fragments into a unique sequence that can be folded into a cloverleaf pattern, consistent with the regular features observed in previously determined sequences? The answer, at least for this particular tRNA, seems to be: almost but not quite.

Possible Nucleotide Sequences of $tRNA_3^{glu}$ of Yeast

The positions of certain fragments are obvious. These include the fragments at the ends (containing a free 3'-OH or phosphate), the fragment containing the anticodon, and the TψC fragment common to all tRNAs. The single fragment containing dihydrouracil was not surprisingly allocated to the dihydroU loop.

We proceeded by trial and error to assemble a sequence that fits the cloverleaf model and most of the other observed regularities. The best sequence found is shown in Figure 1. The positions of two small fragments, a single G and the dinucleotide AG, are not determinate. This structure is in accord with the following regularities of tRNA structure:[1]

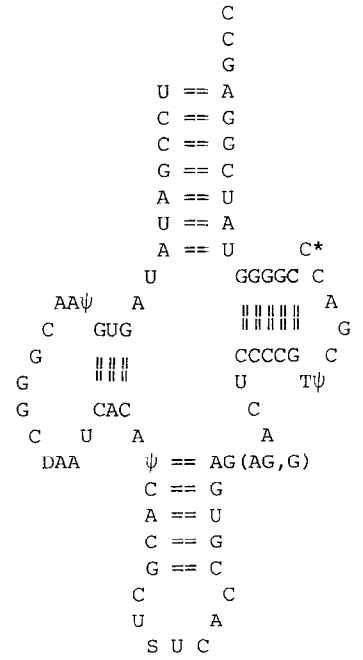

Figure 1. Best structure formed by assembling fragments. (C* = 5-methyl cytosine).

1. The molecule can take up a cloverleaf structure, with the usual numbers of bases in each helical region, and in the loops con-

necting the base-paired regions of the dihydroU arm, the anti-codon arm, and the TψC arm.

2. The first and third base pairs of the dihydroU arm are GC pairs.
3. The base pairs in the TψC arm are GC pairs.
4. Nucleotide 8 is a U.
5. 16 is an A.
6. 17 is a purine.
7. 18 is a pyrimidine.
8. 21 and nucleotide 22 are G's.
9. 31 is a purine.
10. 9 is a purine.
11. 10 is a G.
12. 26 is an A.
13. 63 is a U (complementary to the A at position 17).
14. The sequence of the non-base-paired loop in the TψC arm is of the form T,ψ,C,G,A,anything,pyrimidine.

The structure deviates from observed regularities in certain ways. 1. Position 19 is usually a pyrimidine. However, if positions 18-19-20 are to be pyrimidine-pyrimidine-G, then there must be a T1 fragment ending in this sequence. The possible candidates, fragments 15, 17, 19,

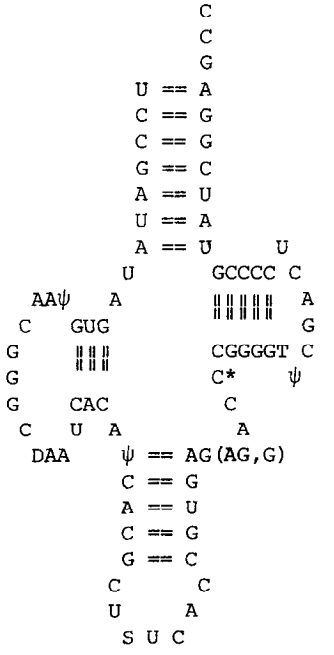

Figure 2. An alternative structure, with the 5-methyl C as position 63, but without complementarity between the bases at positions 17 and 63.

20 and 21, can either be assigned to other parts of the molecule, or contradict the A-purine regularity of positions 16-17. Therefore we feel that position 19 in this molecule is probably not occupied by a pyrimidine.

2. The residue 5-methyl C occurs at an unusual position, in the TψC loop. This base generally occurs at position 63, just before the base-paired region of the TψC arm. An alternative structure in which the 5-methyl C is at position 63 is shown in Figure 2. Here, however, the complementarity between positions 17 and 63 is not present. Therefore, given the information about tRNA structural regularities as it is presently known, the structure in Figure 1 is the more likely. We have not succeeded in finding a sequence that would place the 5-methyl C at position 63 and a G at position 17.

A feature of both structures presented here that has hitherto been unobserved is the occurrence of and AU pair as the central base pair of three in the dihydroU arm. However, GU pairs have been found at this position.

Table 1. T1 fragments of $tRNA_3^{glu}$ from Yeast.[2]

Fragment Number	Sequence
1	CC
2-8	G,G,G,G,G,G,G
9,10	UG,UG
11-13	AG,AG,AG
14	TψCG
15	AC 5-methylC CG
16	ψAACG
17	UAUCG
18	AUAUAG
19	pUCCG
20	ACUCCCCG
21	CUSUCACCG
22	CDAUCACAψCACG

(S = 2-thiouridine-5-acetic acid methyl ester)

Conclusions

In the case of $tRNA_3^{glu}$ of Yeast, knowledge of the sequences of the fragments produced by T1 digestion was sufficient to specify uniquely almost all of the overall nucleotide sequence. The most serious ambiguities could be resolved if it were clear which of the

hypothesized regularities are truly common features of all tRNA species, demanded by structural requisites for biological function, and which are merely coincidences that will not survive in the literature.

Information about tRNA nucleotide sequences is essential for a detailed understanding of their structure, function and evolutionary development. As knowledge of the structural regularities among tRNAs increases, it will become easier to determine additional nucleotide sequences.[3]

Work supported by the National Institutes of Health, U.S. Public Health Service research grants RR-00578 and GM-16539, and U.S. National Science Foundation research grant GB-7857.

References

1. M. Staehelin, Experientia 27, 1 (1971).
2. M. Yoshida, K. Takeishi and T. Ukita, Biochem. and Biophys. Acta 228, 153 (1971).
3. S. Takemura, M. Murakami and M. Miyazaki, J. Biochem. 65, 553 (1969).