THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 256, No. 4, Issue of February 25, pp. 1539-1543, 1981
Printed in U.S.A.

The Evolution of Multi-isoacceptor tRNA Families

SEQUENCE OF $tRNA^{Leu}_{CAA}$ AND $tRNA^{Leu}_{CAG}$ FROM *ANACYSTIS NIDULANS**

(Received for publication, June 6, 1980, and in revised form, September 12, 1980)

Bernard LaRue
From the Département de Chimie-biologie, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada G9A 5H7
Nicole Newhouse, Krikor Nicoghosian, and Robert J. Cedergren
From the Département de Biochimie, Université de Montréal, Montréal, Québec, Canada H3C 3J7
With the technical assistance of David Sankoff
From the Centre des Recherches mathématiques, Université de Montréal, Montréal, Québec, Canada H3C 3J7

Two leucine tRNAs from the cyanophyte *Anacystis nidulans* have been isolated, and their complete nucleotide sequences have been determined by combining data from oligonucleotide fingerprints and sequencing gels. The two sequences are 87 nucleotides long, have the anticodons CAA and CAG, and differ from each other at a total of 28 positions. They have been compared to other known $tRNA^{Leu}$ sequences and incorporated into a phylogenetic tree comprising prokaryotic and chloroplastic $tRNA^{Leu}$ sequences. Mutations inferred from the tree show that some parts of the tRNA molecule are highly variable (the extra arm and the acceptor stem) while others are much more conserved (the D and T arms). The topology of the tree supports the idea that blue-green algae and chloroplasts share a common prokaryotic ancestor and show a basic divergence between XAA and XAG anticodon-containing tRNAs, suggesting that these two subfamilies result from an ancient gene duplication. Finally, comparison of this phylogenetic tree with those of other multi-isoacceptor tRNA families shows no common scheme, which may be due to independent refinement of codon-reading patterns in different tRNA families.

The comparison of structural data has provided useful and sometimes unique insights into the structure-function relationships of the tRNA molecule (1). However, even with the over 150 known tRNA sequences (2), only rather limited evolutionary interpretations have been advanced, since most structural data emanate from such few organisms. Recently, additional sequences from an increasing variety of organisms have kindled the hopes of eventually deducing major evolutionary principles from these data (3). Some years ago, studies on cyanophyte (blue-green algae) tRNA were initiated in our laboratory in order to expand this data set (4, 5). We now report the complete nucleotide sequences of $tRNA^{Leu}_{CAG}$[1] and $tRNA^{Leu}_{CAA}$ from *Anacystis nidulans* and compare them with other leucine tRNA structures in order to rationalize the origin and evolution of this multiisoacceptor tRNA family.

EXPERIMENTAL PROCEDURES[2]

RESULTS AND DISCUSSION

Experimental details of the sequence determination of $tRNA^{Leu}_{CAG}$ and $tRNA^{Leu}_{CAA}$ are given in a supplement which follows this article. Analysis of these data permitted the deduction of the two unique sequences shown in Fig. 1. Both sequences are 87 nucleotides long, a length dictated by an extended extra arm of 15 nucleotides. The anticodon sequence, CAG, of $tRNA^{Leu}_{CAG}$ is followed by m^1G, which is also found at this position in several tRNAs with an anticodon ending in G (2, 6). An unidentified nucleoside, presumably a hypermodified adenosine derivative, is adjacent to the 3′ side of the CAA anticodon of $tRNA^{Leu}_{CAA}$. The two sequences differ from each other at a total of 28 positions which, except for the identical D arm, are scattered over the entire tRNA molecule. The conservation of the D arm in the two tRNAs is consistent with its known involvement in tertiary structure interactions (1). Both leucine tRNAs can be written in the usual cloverleaf conformation (Fig. 1), with nonstandard base pairs occurring only at the loop end of the D stem and in the extra arm. In addition, the leucine tRNAs possess all 23 so-called invariant and semi-invariant residues (1). $tRNA^{Leu}_{CAG}$ and $tRNA^{Leu}_{CAA}$ contain seven and six modified nucleosides, respectively, which consist primarily of the usual uridine derivatives, D, T, and Ψ. Such low modification levels are also typical of the methionine initiator and phenylalanine tRNAs from *A. nidulans* (5, 2). As in other tRNAs having a long extra arm, a nonstandard base pair between positions 13 and 22 and a $purine_{26}$-$pyrimidine_{44}$ combination are present (8).

In order to assess the evolutionary significance of these algal tRNAs within the leucine multi-isoacceptor tRNA family, we have compared all 17 known sequences, that is, the 2

* This study was supported by the Medical Research Council of Canada and the Faculté de Médecine of the Université de Montréal. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] The abbreviations used are: $tRNA^{AA}_{XXX}$, tRNA specific for the amino acid AA and having the anticodon XXX; m^1G, 1-methylguanosine; U*, unknown modification of uridine; N, unknown modification; *A, unknown modification of adenosine; X, any of the four common nucleosides. Fingerprint and gel sequencing data have been deposited at the *Journal of Biological Chemistry* office.

[2] Portions of this paper (including "Experimental Procedure," part of "Results," Figs. 3 and 4, and Table II) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, Md. 20014. Request Document No. 80M-1141, cite authors, and include a check or money order for $2.80 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

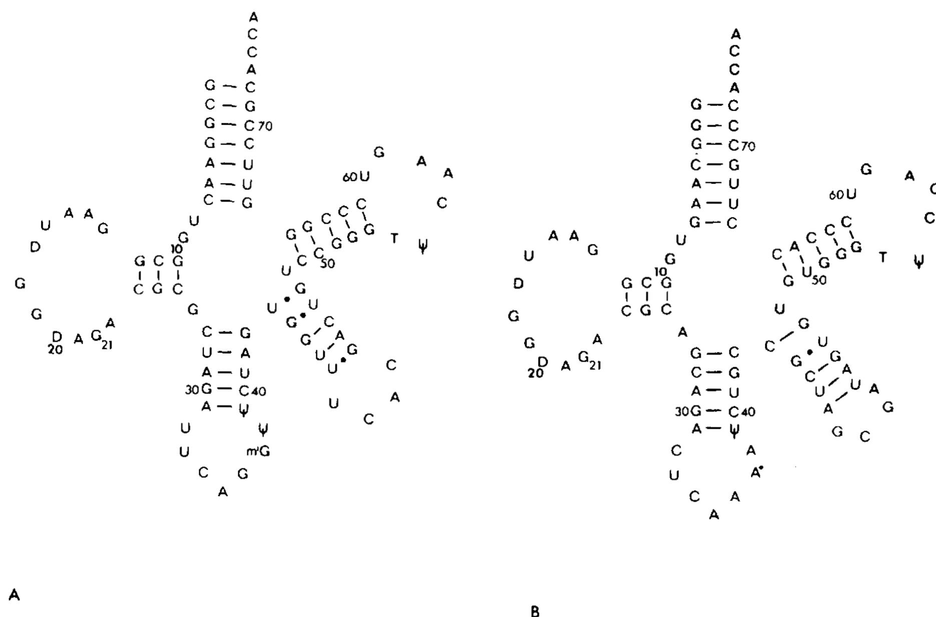

FIG. 1. **Cloverleaf structures of tRNA$^{Leu}_{CAG}$ (*A*) and tRNA$^{Leu}_{CAA}$ (*B*).** *Numbered positions* refer to the Cold Spring Harbor numbering system (2).

TABLE I
Difference matrix of sequenced leucine tRNAs

Each sequence is identified according to species and anticodon. Positions of extra arm that are unfilled in most other types of tRNAs and the anticodon nucleotides were not considered. Differences were counted after conversion of modified nucleosides to their parents.

Sequence number	Species anticodon	Mitochondria				Eukaryotes						Prokaryotes				
		Sequence number														
		16	15	14	13	12	11	10	9	8	7	6	5	4	3	2
1	T$_4$ UAA	35	44	39	33	37	36	26	30	25	34	27	28	24	23	24
2	*E. coli* AAA	25	37	42	38	44	40	36	25	30	25	26	27	23	26	
3	*A. nidulans* CAA	28	41	37	32	35	34	38	28	33	31	24	21	17		
4	Bean chloroplast UAA	29	37	37	33	39	32	40	26	27	31	28	24			
5	*A. nidulans* CAG	38	32	37	31	36	32	34	30	26	17	15				
													Prok-prok[a]			
6	*E. coli* CAG	36	34	38	30	38	34	30	28	30	16					
7	*E. coli* GAG	32	20	42	39	41	40	34	26	22						
8	Bean chloroplast UAG	36	31	44	44	36	39	32	25							
9	Bean chloroplast CAA	30	37	40	38	38	42	40								
								Euk-prok[a]								
10	*Xenopus* CAG	40	43	28	28	28	32									
11	*S. pombe* UAA	42	40	25	21	29										
						Euk-euk[a]										
12	*T. utilis* CAA	52	41	13	21											
13	*S. cerevisiae* UAG	43	43	14												
14	*S. cerevisiae* CAA	48	44													
			mt-other[a]													
15	*N. crassa* mt UAG	36														
16	*N. crassa* mt UAA															

[a] Prok-prok indicates comparisons among prokaryotic sequences, and euk-prok and euk-euk indicate differences among eukaryotic and either prokaryotic or eukaryotic structures. mt refers to mitochondrial comparisons.

sequences from *A. nidulans*, 4 from *Escherichia coli* (2), 1 from the bacteriophage T$_4$, 3 from bean chloroplast (9), 2 from *Neurospora crassa* mitochondria (10), and 5 from various eukaryotic cytoplasms (2). The number of structural differences between each pair of tRNA^{Leu} is shown in Table I. The difference matrix is calculated after modified nucleosides are converted to their parent and does not take into account differences in generally unfilled positions of the variable loop (47.1–47.16) and in the anticodon. This comparison permits the observation that the average number of structural differences among prokaryotic tRNAs (including the chloroplastic sequences) and among eukaryotic tRNAs (25.6 and 23.9, respectively) is significantly lower than the average difference between a eukaryotic and a prokaryotic tRNA (36.4).

As shown in Fig. 2, we have constructed a minimal mutation phylogenetic tree (11) with the prokaryotic tRNA^{Leu} sequences. However, we have excluded from this sample the chloroplastic tRNA$^{Leu}_{CAA}$, since it is quite distant from all other prokaryotic sequences (see Table I), and it is not possible, using the presently known sequences, to find for it a clearly optimal position in the tree. We have identified the 116 mutations resulting from the topology of the tree and have found 41 additional mutations when the entire extra loop region is included. The D, T, and anticodon loops represent 25% of the sequence, but only 8% of the mutations; this evolutionary conservatism correlates well with the importance of these regions in tRNA structure and function (1). The relative frequencies of mutations expressed as the value per

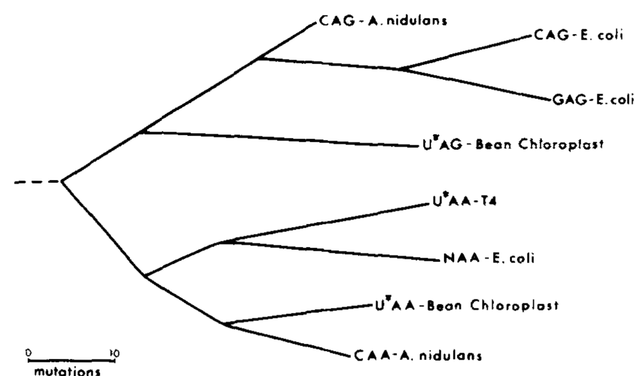

FIG. 2. **Phylogenetic tree of prokaryotic leucine tRNAs.** The tree has been derived according to a minimal mutation criterion (11) and the horizontal distance separating two adjacent nodes has been set proportional to the number of mutations intervening between them. The anticodon nucleotides and the variable length part of the extra arm were excluded from the procedure. The root divides the tree into two equal halves, considering the weighted contribution of each branch.

position normalized to the D arm are the following: D arm, 1.0; acceptor stem, 3.2; T arm, 1.4; anticodon arm, 2.7; and the variable part of the extra arm, 4.9. This pattern is quite similar to the one found in other tRNA families (3), suggesting again that identical structural constraints are present in all types of tRNAs (perhaps excluding mitochondrial tRNAs). Interestingly, the extra arm is so variable that there is usually no more similarity in this region between two tRNAs than between random sequences. Therefore, it is likely that its sequence as such is of little functional significance, although arm length may be an important factor.

If we assume a more or less constant mutation rate among prokaryotic and chloroplastic leucine tRNAs, a root located in the middle of the phylogenetic tree, such as in Fig. 2, would divide it between the XAA and XAG anticodon-containing tRNAs. This division according to codon specificity, either CUX or UUA/G, would support the idea that a minimum of two different leucine tRNAs was required from an early evolutionary date in order to read all six codons from the two coding groups (12). The sparsity of data, however, does not allow us to determine the chronology of this divergence compared to other early events such as the eukaryotic-prokaryotic split. Still, we notice that the mitochondrial sequences (although too distant to be usable at the present time) are consistently closer to the prokaryotic sequences of the same anticodon (XAA or XAG) family (see Table I). This fact gives credence to the hypothesis not only of a very early separation of these two isoaccepting subfamilies, but also of the eventual utility of the mitochondrial tRNA sequences in evaluating evolutionary events.

No clear-cut division according to anticodon type is obvious within each of the two subfamilies. On the contrary, both speciation and gene duplication events seem to occur at various levels of each subfamily. Looking first at the XAG anticodon family, the 15 differences found between *A. nidulans* and *E. coli* $tRNA^{Leu}_{CAG}$ compare quite well with the 12 differences between phenylalanine tRNAs from the same organisms (13). Consequently, they should reflect the divergence between the *E. coli* and blue-green algal lines; on the other hand, the *E. coli* line is further split into $tRNA^{Leu}_{CAG}$ and $tRNA^{Leu}_{GAG}$, showing a more recent gene duplication (Fig. 2). Also, while much evidence (*e.g.* Ref. 13) points to a common origin of chloroplasts and blue-green algae, the tree of Fig. 2 and the large distance between the chloroplastic $tRNA^{Leu}_{U^*AG}$ and *A. nidulans* $tRNA^{Leu}_{CAG}$ indicate, in this case, an ancient gene duplication rather than speciation.

In the XAA anticodon family, we note that $tRNA^{Leu}_{U^*AA}$ from T_4 and $tRNA^{Leu}_{NAA}$ from *E. coli*, although distantly related, are found on the same branch. Previously, it was shown with $tRNA^{Gly}$ sequences that the most recent ancestor of phage and *E. coli* existed after the divergence of the *Bacilli* (13) and, in this sense, the leucine tRNA tree is compatible with this earlier result. Fig. 2 shows a rather close relationship between *A. nidulans* $tRNA^{Leu}_{CAA}$ and the chloroplastic $tRNA^{Leu}_{U^*AA}$; also, the number of differences involved (17 differences) is of comparable magnitude to the number between *A. nidulans* and chloroplastic $tRNA^{Phe}$ sequences (13). Thus, tRNA sequence data support the hypothesis of a blue-green algae-chloroplast common ancestry (14).

In summary, it appears at best hazardous to use a collection of tRNA sequences with a unique anticodon, but belonging to a multi-isoacceptor family, in order to obtain a phylogeny of the species involved. Indeed, each of the three other such families studied so far, those of the glycine (13), lysine, and eukaryotic serine (3) tRNAs, has a unique evolutionary pattern which comprises both gene duplication and speciation events. On the other hand, the phylogenies obtained from the methionine initiator and phenylalanine tRNA families (15), each possessing only one anticodon type, reproduce those based on protein or ribosomal RNA sequence data. The peculiar behavior of multi-isoacceptor tRNA families is still understandable in terms of the "two out of three" hypothesis (12), which states that the third (3′) codon position has little, if any, discriminating function in several codon families. Correspondingly, if a given tRNA is able to recognize a whole set of codons differing only at the third position, a duplication of its gene followed by a mutation at the first (5′) anticodon position would pose no threat to the accuracy of the translation process. Such a mechanism would explain why anticodon changes take place in multi-isoacceptor tRNA families, and it is hoped that a study of the frequency and nature of these changes will clarify the constraints responsible for the efficiency and the selectivity of the translation apparatus.

REFERENCES

1. Rich, A., and RajBhandary, U. L. (1976) *Annu. Rev. Biochem.* **45,** 805–860
2. Sprinzl, M., Grueter, F., Spelzhaus, A., and Gauss, D. H. (1980) *Nucleic Acids Res.* **8,** r1–r22
3. Cedergren, R. J., LaRue, B., Sankoff, D., and Grosjean, H. (1980) *CRC Crit. Rev. Biochem.* in press
4. Beauchemin, N., LaRue, B., and Cedergren, R. J. (1973) *Arch. Biochem. Biophys.* **156,** 17–25
5. Ecarot-Charrier, B., and Cedergren, R. J. (1976) *FEBS Lett.* **63,** 287–290
6. Nishimura, S. (1972) *Prog. Nucleic Acid Res. Mol. Biol.* **12,** 49–85
7. Deleted in proof.
8. Brennan, P., and Sundaralingam, M. (1976) *Nucleic Acids Res.* **3,** 3235–3251
9. Osorio-Almeida, M. L., Guillemaut, P., Keith, G., Canaday, J., and Weil, J. H. (1980) *Biochem. Biophys. Res. Commun.* **92,** 102–108
10. Heckman, J. E., Sarnoff, J., Alzner-Deweerd, B., Yin, S., and RajBhandary, U. L. (1980) *Proc. Natl. Acad. Sci. U.. S. A.* **77,** 3159–3163
11. Moore, G. W., Barnabas, J.., and Goodman, M. (1973) *J. Theor. Biol.* **38,** 459–485
12. Lagerkvist, U. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 1759–1762
13. Cedergren, R. J., LaRue, B., Sankoff, D., Lapalme, G., and Grosjean, H. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2791–2795
14. Margulis, L. (1970) *Origin of Eukaryotic Cells,* Yale University Press, New Haven, Conn.
15. LaRue, B., Cedergren, R. J., Sankoff, D., and Grosjean, H. (1979) *J. Mol. Evol.* **14,** 287–300

16. Ecarot, B., and Cedergren, R. J. (1974) *Biochem. Biophys. Res. Commun.* **59**, 400–405
17. Gillam, I., Blew, D., Warrington, R. C., von Tigerstrom, M., and Tener, G. M. (1968) *Biochemistry* **7**, 3459–3468
18. Pearson, R. L., Weiss, J. F., and Kelmers, A. D. (1971) *Biochim. Biophys.. Acta* **228**, 770–774
19. Maxam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. **74**, 560–564
20. Brownlee, G. G. (1972) in *Determination of Sequences in RNA*, American Elsevier, New York
21. Bruce, A. G., and Uhlenbeck, O. C. (1978) *Nucleic Acids Res.* **5**, 3665–3677
22. Peattie, D. A. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1760–1764
23. Wurst, R. M., Vournakis, J. N., and Maxam, A. M. (1978) *Biochemistry* **17**, 4493–4499
24. Fournier, M. J., Webb, E., and Kitchingman, G. R. (1976) *Biochim. Biophys. Acta* **454**, 97–113

SUPPLEMENT

The Evolution of Multi-isoacceptor tRNA Families : Sequence of $tRNA^{Leu}_{CAA}$ and $tRNA^{Leu}_{CAG}$ from Anacystis nidulans.

Bernard LaRue, Nicole Newhouse, Krikor Nicoghosian and R. J. Cedergren

EXPERIMENTAL PROCEDURES

Anacystis nidulans (strain 625, Algal Culture Collection of the University of Indiana) was grown and RNA was extracted as previously described (4). Radioactive tRNA was obtained by growing cells in the presence of 50 mCi of carrier-free $[^{32}P]H_3PO_4$ from New England Nuclear (16). The labelled cells were subjected to phenol extraction and RNA was precipitated after the addition of 1 mg of non-radioactive algal tRNA. The RNA was dissolved in 2 ml of 0.1 M NaCl-0.01 M sodium acetate-0.01 M $MgCl_2$ (pH 4.5) and adsorbed onto a 2-ml volume BD-cellulose column. The tRNA fraction is that which elutes between 0.3 M NaCl and 1 M NaCl-20% ethanol in the same acetate-magnesium buffer. After ethanol precipitation, the tRNA was deacylated by a 30 min, 37° treatment in 1 ml of 1 M Tris-HCl (pH 9.0) containing 2 mM β-mercaptoethanol. The tRNA was precipitated and aminoacylated in 2.6 ml of a 0.05 M Tris-H_2SO_4 (pH 7.5) buffer containing 92.5 μmoles of $MgCl_2$, 7.5 μmoles of ATP, 7.5 μmoles of CTP, 0.08 μmole of $[^3H]$ leucine (10 Ci/mmole, New England Nuclear) and 5.6 mg of E.coli aminoacyl-tRNA synthetases prepared as described previously (4). After incubation for 30 min at 37°, 1 ml of phenol was added, and the aminoacylated tRNA was precipitated from the aqueous layer by the addition of 3 volumes of ethanol. Phenoxyacetylation was accomplished at 0° by adding 10 mg of the phenoxyacetyl ester of N-hydroxysuccinimide in 0.1 ml of dry tetrahydrofuran to the tRNA dissolved in 1 ml of 0.1 M triethanolamine at pH 8.0 (17). The reaction was stoped by the addition of 5 μl of acetic acid. After centrifugation at 10,000xg for 10 min, the supernatant was diluted with 3 ml of a 0.01 M sodium acetate (pH 4.5) buffer containing 0.3 M NaCl and 0.01 M $MgCl_2$.

The tRNA was adsorbed on a 4-ml volume BD-cellulose column, which was then extensively washed with the above buffer containing 1 M NaCl. Phenoxyacetyl-$[^3H]$leucyl-$[^{32}P]$ tRNA was eluted in the acetate-magnesium buffer containing 1.0 M NaCl and 20% ethanol. The ethanol fraction (approx. 12 ml) was collected and dialyzed at 4° for 1 hr against 1.0 l of the acetate-magnesium buffer. This solution was run onto a 0.4 x 100 cm column containing RPC-5 packing (18) which was developed with a 300-ml linear gradient of 0.5 to 1.5 M NaCl in the same buffer. Two of the major peaks, I (leucine tRNA with a CAG anticodon) and III (leucine tRNA, CAA anticodon) were separately pooled, deacylated and dialyzed for 2 hours as above. Each tRNA solution was then adsorbed onto a second RPC-5 column (0.4 x 100 cm) which was eluted by a 200-ml linear gradient of 0.4 M to 1 M NaCl in the acetate magnesium buffer. The radioactive peak from each tRNA sample was concentrated by adsorbing on a 1-ml BD-cellulose column and eluting with a 0.05 M Tris-HCl (pH 7.5) buffer containing 0.01 M $MgCl_2$, 1 M NaCl and 20% ethanol. A sample of 50 μg of carrier tRNA was added to the recovered radioactive fraction, and precipitation of the tRNA was accomplished by the addition of 2 volumes of ethanol.

Non-radioactive leucine tRNAs used for in vitro labelling studies were purified by an alternate procedure. For aminoacylation, a sample of 25 mg of deacylated algal tRNA was incubated for 30 min at 37° in 20 ml of a 0.05 M Hepes buffer (pH 7.5) containing : 1.28 mmoles of $MgCl_2$, 0.16 mmole of ATP, 0.16 mmole of CTP, 0.122 μmole of $[^{14}C]$ leucine (287 mCi/mmole, New England Nuclear) and 26 mg of E.coli synthetases prepared as in Ref. 4, concentrated and stored in glycerol. Phenoxyacetylation was carried out as above using 3 ml of triethanolamine buffer to dissolve the tRNA and 50 mg of the phenoxyacetyl ester in 300 μl of tetrahydrofuran. The phenoxyacetylated product was then applied to a 10 ml column of BD-cellulose; after removal of the 1 M NaCl fraction, the radioactive fraction was eluted by the above acetate-magnesium buffer containing 1 M NaCl and 20% ethanol. This 15 ml fraction was dialyzed against 2 l of the acetate-magnesium buffer and run onto a RPC-5 column, which was eluted with the same gradient detailed above. Two of the major radioactive peaks corresponding to the same peaks I and III of the $[^{32}P]$ sample were pooled, concentrated to 0.3 ml on a Diaflo UM-10 membrane and ethanol precipitated. Each tRNA preparation was taken up in 100 μl of 0.25 M Tris-acetic acid (pH 9.0) incubated at 37° for 45 min and ethanol precipitated. The deacylated tRNA was dissolved in 10 μl of 0.05 M Tris-boric acid (pH 8.3) containing 7 M urea and 1 mM EDTA and subjected to 15% polyacrylamide-7M urea gel electrophoresis (slab 20 x 40 cm) at 600 volts (19). The cyanol blue marker was allowed to migrate 38 cm. The tRNA band was located by illumination of the slab with a UV light source, cut from the slab and extracted by the overnight incubation of the ground acrylamide in 300 μl of 0.5 M ammonium acetate, 0.01 M magnesium acetate, 0.1% sodium dodecylsulfate and 0.1 mM EDTA. The tRNA was precipitated from the filtered solution by the addition of 2.5 volumes of ethanol The amino acid acceptance activity was determined as above.

In vivo labelled tRNAs were digested with pancreatic or T_1 ribonuclease, and products were separated by two-dimensional electrophoresis on cellulose acetate strips (Schleicher and Schuell) at pH 3.5 and DEAE-paper in 7% formic acid following the technique published by Brownlee (20). The nucleotide composition of individual oligonucleotides isolated from the above digests were obtained by digestion with T_2 ribonuclease, and further structural information was deduced from secondary enzymatic digests using pancreatic, T_1 and U_2 ribonuclease and snake venom phosphodiesterase. Oligonucleotides were also systematically analyzed for modified nucleosides by two-dimensional thin layer chromatography on cellulose plates after T_2 ribonuclease digestion (6).

In vitro labelled tRNA was prepared by the method of Bruce and Uhlenbeck using 0.6 μg of tRNA, $[^{32}P]pCp$ (1900 Ci/mmole, New England Nuclear) and RNA ligase from P-L Biochemicals (21). The resulting labelled RNA was purified by electrophoresis on 20 x 20 cm slabs of 15% polyacrylamide gels as above; RNA was extracted from the gel and processed as in the chemical degradation sequencing technique of Peattie (22) using the following modifications : the guanosine-specific degradation was performed with 1.5 μl of dimethyl sulfate for 1.5 min, the adenosine reaction with diethyl pyrocarbonate was for 3 min and the cytidine and uridine reactions were for 8 min. The 3'-labelled RNA fragments were separated on a 15% polyacrylamide slab gel (20 x 52 x 0.08 cm) in 0.05 M Tris-boric acid containing 7 M urea and 1 mM EDTA. Gels were run at 1500-2000 volts. In other experiments, tRNA, following dephosphorylation, was 5'-labelled using polynucleotide kinase (Boehringer-Mannheim) and γ-$[^{32}P]$ATP (3300 Ci/mmole, New England Nuclear) as described by Wurst et al. (23). Sequence analysis of this material was performed as above.

The number of differences between tRNA pairs and the mutations inferred in phylogenetic trees were calculated as described previously (3, 13). The tRNA sequence data base comes from the compilation of Sprinzl et al. (2). The chloroplastic and mitochondrial leucine tRNA sequences were graciously provided prior to publication by Professors J.H. Weil and U.L. RajBhandary, respectively.

RESULTS

The aminoacylation of both labelled and unlabelled leucine tRNA was effected with a crude E. coli aminoacyl-tRNA synthetase preparation, since it had previously been shown that the E. coli enzyme fully recognizes all five algal leucine tRNAs (4). Following phenoxyacetylation, the leucyl-tRNAs were first eluted in bulk as the ethanol fraction from a BD-cellulose column and next resolved into three peaks on a RPC-5 column (Fig. 3). In the final purification step of the in vivo labelled tRNAs, we have taken advantage of the much earlier elution from a RPC-5 column (0.3 - 0.5 M lower NaCl concentration) of the deacylated tRNA as compared to its phenoxyacetylated counterpart. Peak II from Fig. 3 was shown by subsequent analysis to be a mixture of two leucine isoacceptor tRNAs and was not further characterized. Peaks I and III correspond to the leucine tRNA (CAG anticodon) and leucine tRNA (CAA anticodon) respectively. After the second RPC-5 chromatography, fingerprint analysis of leucine tRNA-CAG and leucine tRNA-CAA showed the purity of these tRNAs to be generally greater than 90%. The purification of the non-radioactive leucine tRNAs followed the same scheme except that a 15% polyacrylamide gel electrophoresis step was substituted for the second RPC-5 chromatography. The specific activities of Leu tRNA-CAG and Leu tRNA-CAA prepared by this method were 1290 pmoles/A_{260} and 920 pmoles/A_{260} respectively. These rather low acceptor activities may be gel purification artefacts, since the subsequent sequencing gels do not show heterogeneity.

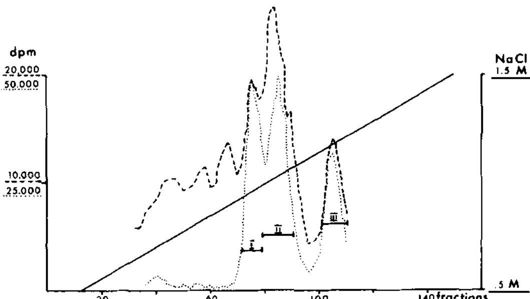

Fig. 3 Elution pattern from a RPC-5 column of phenoxyacetyl-$[^3H]$ leucyl $[^{32}P]$ tRNAs. The tRNA material eluted in the ethanol fraction of a BD-cellulose column was applied to a 0.4 x 100 cm RPC-5 column which was developed by a 300-ml linear gradient (0.5-1.5 M NaCl) in 0.01 M sodium acetate-0.01 M $MgCl_2$ (pH 4.5). ^{32}P --------, ^3H ....

The oligonucleotides produced by T_1 and pancreatic RNase digestion of the two tRNAs were separated by two-dimensional electrophoresis and characterized as described in Experimental Procedures (see footnote to the Supplement). The sequence of the oligonucleotides is given along the linearly written sequences of the two tRNAs shown in Fig. 4. The molarity of all oligonucleotides in the fingerprints is within 20% of their theoretical value except for the 3' and 5' terminal oligonucleotides, which were usually obtained in lower yield than expected. In the pancreatic RNase digest of Leucine tRNA-CAG, spots 17 and 18 correspond to equimolar amounts of GGU/GGD and GGGU/GGGT respectively as determined by thin layer chromatography. In the T_1 RNase digest of Leucine tRNA-CAA, we note that spot five is a mixture of two moles of CAG and one mole of ACG. In addition, spot 13 from the same fingerprint is a mixture of two oligonucleotides which could not be separated. The sequence of these two oligonucleotides could not be deduced solely from the in vivo labelling experiments (see below).

The arrangement of oligonucleotides in the two tRNAs shown in Fig. 4 was determined by the gel separation of oligonucleotides produced by chemical degradation procedures (22). The sequencing gels show artefacts which would make it very difficult to determine the whole sequence by using this technique alone; the 5' stem region of the T loop is subject to serious band compacting and the doublet nature of certain bands is especially evident near the 3' terminal region of both gels. A further problem with this technique is the identification of modified nucleosides. However, the identity of some of these can be inferred from anomalous gel patterns. The behaviour of some nucleosides which are present in the two leucine tRNA sequences and other tRNAs recently studied in our laboratory are indicated in Table II.

TABLE II : BEHAVIOUR OF NUCLEOSIDES ON SEQUENCING GELS.

Name	Symbol	Row of the sequencing gel G	A	C	U
Thymidine	T	-	-	-	-
Pseudouridine	ψ	-	-	-	-
Dihydrouridine	D	-	-	++	+++
7-Methylguanosine	m^7G	+++	-	+++	+++
1-Methylguanosine	m^1G	++	+	-	-
5-Methylcytidine	m^5C	-	-	-	-
1-Methyladenosine	m^1A	-	-	-	-
"Y" base	Y	+	+++	-	-
2-methylthio-6-isopentenyl adenosine	ms^2i^6A	-	+++	-	-
unknown adenosine derivative	*A	-	+++	+	-
guanosine	G	++	+	-	-
adenosine	A	+	++	-	-
cytidine	C	-	-	++	+
uridine	U	-	-	+	+++

Legend :
+++ : very dark ++ : dark +: faint - : very faint or absent band

Although neither sequencing method by itself would have provided the complete sequence, the combination of the two, in vivo and in vitro, labelling techniques permitted the elucidation of an unique sequence. Only one discrepancy is apparent in the two data sets. The identity of position 40 in the Leu tRNA-CAG is ψ. The nucleotide of position 40 in Leu tRNA-CAG was identified as a U after analysis of spot 18 of the pancreatic RNase digest. The gel, however, shows this nucleotide to be a pseudouridine, since there is no band evident in the U track (Table II). The undermodification of the in vivo-labelled tRNA could be due to a nutritional deficiency resulting from the low phosphate concentration used during $[^{32}P]$ labelling (24). In addition, results from the two techniques allowed the correct deduction of the two oligonucleotide sequences of the T_1 RNase-generated spot 13 from Leucine tRNA-CAA. Here although secondary digests of the mixture with pancreatic, U_2 and T_2 RNases had indicated the presence of AA·AAψ, AC, C_2U_2G and CU_2G, there was no information on the relative position of these oligonucleotides. Finally, the use of a 5' labelled Leucine tRNA-CAA sample followed by chemical degradation confirmed the sequence of the first 15 nucleotides : $_p$GGGCAAGUGGCGGAA.

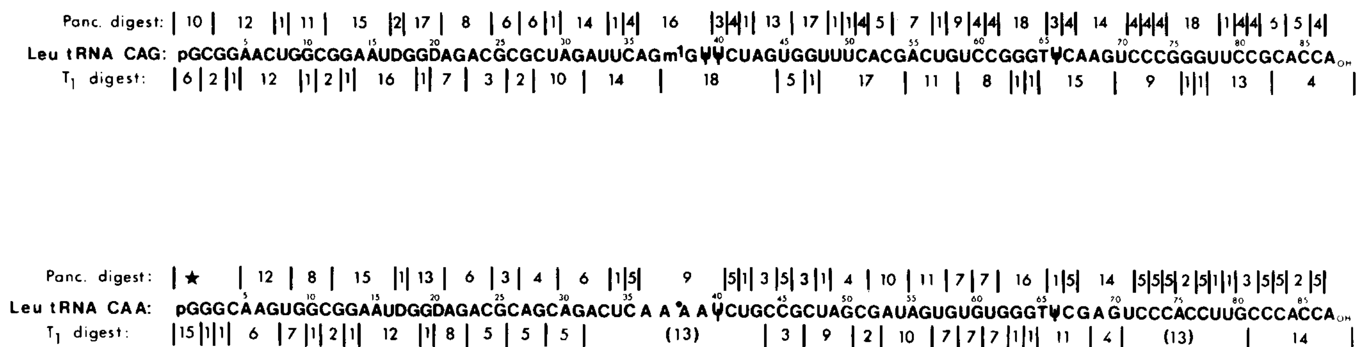

FIG. 4. Linear sequences of leucine tRNAs and identification of oligonucleotides. Both sequences are written linearly and numbers between vertical bars correspond to spot numbers in their respective fingerprints. The number in parenthesis, (13), was not deduced through the fingerprinting method; *, oligonucleotide not found in the fingerprint. °A: unidentified adenosine derivative.