J. Mol. Evol. 14, 287–300 (1979)

Journal of Molecular Evolution

Evolution of Methionine Initiator and Phenylalanine Transfer RNAs

B. Larue,[1]* R.J. Cedergren, D. Sankoff,[2] and H. Grosjean[3]

[1] Département de Chimie-Biologie, Université du Québec à Trois-Rivières, Case postale 500, Trois-Rivières, Québec, Canada;
[2] Département de Biochimie et Centre de Recherches Mathématiques, Université de Montréal, Case postale 6128, Montréal, Québec, Canada;
[3] Laboratoire de Chimie Biologique, Rue des Chevaux, 67, 1640 Rhode St-Genèse, Belgique

Summary. Sequence data from methionine initiator and phenylalanine transfer RNAs were used to construct phylogenetic trees by the maximum parsimony method. Although eukaryotes, prokaryotes and chloroplasts appear related to a common ancestor, no firm conclusion can be drawn at this time about mitochondrial-coded transfer RNAs. tRNA evolution is not appropriately described by random hit models, since the various regions of the molecule differ sharply in their mutational fixation rates. 'Hot' mutational spots are identified in the TψC, the amino acceptor and the upper anticodon stems; the D arm and the loop areas on the other hand are highly conserved. Crucial tertiary interactions are thus essentially preserved while most of the double helical domain undergoes base pair interchange. Transitions are about half as costly as transversions, suggesting that base pair interchanges proceed mostly through G-U and A-C intermediates. There is a preponderance of replacements starting from G and C but this bias appears to follow the high G + C content of the easily mutated base paired regions.

Key words: Phenylalanine tRNA – Methionine initiator tRNA – Evolution – Mutations – Conformation

Introduction

The complete sequence of over 120 transfer RNAs are currently availabe (Gauss et al., 1979). This data set is rather unevenly distributed among the tRNAs specific for the 20 amino acids and their 61 codons of the genetic code; however, the sample does cover a wide range of prokaryotic and eukaryotic organisms as well as some organelles. In spite of the very similar secondary (cloverleaf) and tertiary structures of tRNAs (Holley et al., 1965; Rich and RajBhandary, 1976; Schewitz et al., 1979) no consistent nor

* To whom offprint requests should be sent

0022–2844/79/0014/0287/$ 02.80

convincing phylogenetic picture has appeared comprising all tRNAs (Schwartz et al., 1976; Hasegawa, 1978). This situation is due to a number of factors such as: short sequence lengths of tRNA, use of only four nucleotides, cloverleaf-imposed base pair restrictions, divergent histories of tRNA isoacceptors (tRNAs which can be aminoacylated by the same amino acid, but possess different codon recognition patterns) and the evolutionary conservatism of a large part of the tRNA molecule. The combination of these factors often obscures evolutionary relationships, since occassionally a given tRNA is quantitatively more related to a tRNA specific for a completely different amino acid than to a tRNA of like activity isolated from another organism (Holmquist et al., 1973; Cedergren et al., 1979).

Therefore, it is crucial to somehow evaluate the true relatedness of distant sets of tRNA data before attempting to derive phylogenetic relationships by classical methods. We recently described such a method based on the convergence of independently derived ancestral sequences and used it to demonstrate that methionine initiator and phenylalanine tRNAs, excluding mitochondrial sequences, each constitute a homologous family (Cedergren, et al., 1979).

We present here an in depth analysis of the evolutionary histories of these two tRNA families made possible by the large number of known sequences available as well as the fact that these two families are composed of monofunctional tRNA species having identical codon recognition patterns.

Sequence Alignment and Phylogenic Tree Derivation

All the sequences studied are tabulated in the review of Gauss et al. (1979). We include five additional sequences kindly provided by the authors prior to publication: $tRNA^{Phe}$ from *Xenopus laevis* (Clarkson, personal communication), $tRNA^{Phe}$ from *S. cerevisiae* mitochondria (Martin et al., 1978), $tRNA^{fMet}$[1] from *Scenedesmus obliquus* (Olins, personal communication), $tRNA^{fMet}$ from *Streptococcus fecalis* (Rabinowitz and Delk, personal communication), $tRNA^{fMet}$ from *Thermus thermophilus* (Oshima, personal communication) and $tRNA^{fMet}$ from *Drosophilia melanogaster* (Silverman et al., 1979).

When comparing sequences, modified nucleosides were first converted to their parent base U, C, A or G. Sequences were aligned according to common structural elements: base-paired regions of the cloverleaf, anticodon, TψC loop, CCA terminus and other invariant residues; matches were maximized in variable length regions by the introduction of gaps. The position numbering system is that agreed upon at the Cold Spring Harbor meeting on tRNA in 1978 and described by Gauss et al. (1979).

Phylogenetic trees were constructed to satisfy as far as possible a maximum parsimony criterion (Moore et al., 1973; Dayhoff et al., 1972a; Dayhoff et al., 1972b) (Fig. 1 and Fig. 2). This involves two minimizations, one embedded in the other. For a given tree topology (branching sequence), the evolutionary history which implies a minimal number of mutations is derived for each position in the sequence. This is carried out for a wide range of possible topologies and the one with the minimal total number of mutations is selected as the most likely to represent the true evolutionary history. In

---

[1] $tRNA^{fMet}$: Symbol used for initiator methionine tRNA.

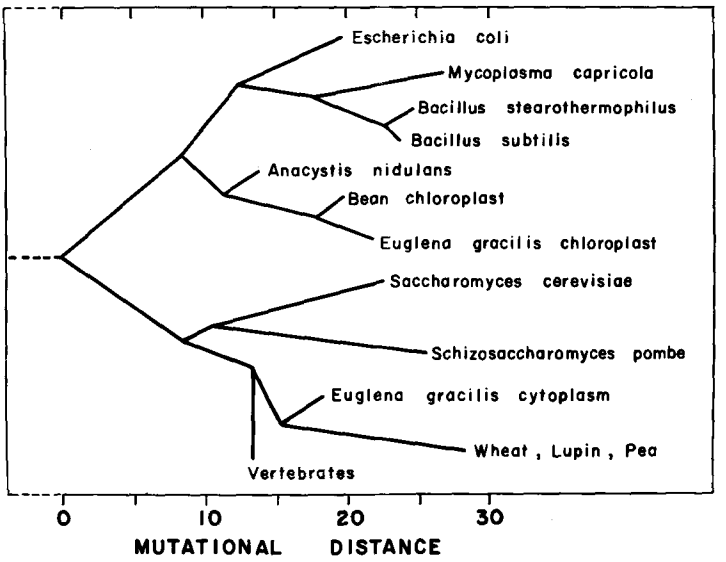

Fig. 1. Phylogeny of $tRNA^{Phe}$, excluding the mitochondrial sequence. Origin of the tree was arbitrarily set at an equal distance from the prokaryotic and eukaryotic roots. Each type of replacement is assigned a uniform weight of 1; fractional values arise from ancestral sequences uncertainties. Distance between a node and its nearest neighbor is given by abscissa diffference value. This figure is based on 12 sequences differing by more than two positions

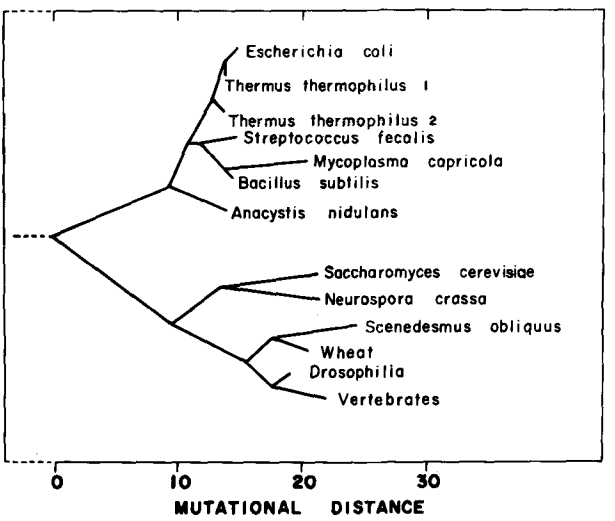

Fig. 2. Phylogeny of $tRNA^{fMet}$, excluding the mitochondrial sequence. Legend as in Fig. 1. The figure is based on 13 sequences

doing this, each inferred point mutation was assigned a uniform cost of 1, since the probabilities of the various mucleotide interchanges are not known for tRNA. One exception to the minimal tree concerns the $tRNA^{Phe}$ sequence from *Euglena*. While this tRNA would cluster with vertebrates using the minimum mutation model, consideration of ancestral sequence convergence between eukaryotes and prokaryotes clearly indicates that Euglena branches either with the plants or prior to the plant-animal divergence (Cedergren, et al., 1979). As a final comment on tRNA phylogenetic trees, we find that in fact many alternate prokaryotic topologies give very similar mutation data, and it is difficult to establish that one of these trees is unambiguously better than another without using some biological criterion. On the other hand the results in this paper should be considered valid, since the mutational patterns described herewithin are rather insensitive to the location of the root of the tree or alternate prokaryotic phylogenies. For the sake of consistency, the trees selected here are the ones with the minimum total mutational distance when similar branching orders are used for both the $tRNA^{Phe}$ and the $tRNA^{fMet}$ families.

Organelle tRNAs

Although a clear phylogenetic relationship can be shown for eukaryotic and prokaryotic members of the initiator and phenylalanine tRNA families, there are little statistical grounds for placing mitochondrial sequences from *N. crassa* or *S. cerevisiae* at any site in these trees[2]. Although mitochondrial tRNAs could have independent origins, that is, derived from ancestors with different acceptors activities, this is unlikely to be true in the three cases, since it would involve a major reshuffling of tRNAs with respect to the genetic code. Nevertheless an additional series of oddities set mitochondrial tRNAs apart:

– As already noticed by others (Dirheimer et al., 1979), the base paired regions of the cloverleaf have an unusually low G + C content.

– In *N. crassa* mitochondrial $tRNA^{fMet}$, there is a standard base pair at the end of the acceptor stem, a eukaryotic trait, but the TψC purine (prokaryotic) or AUCG (eukaryotic) sequence of the TψC arm is replaced by UGCA (Heckman et al., 1978). Of the seven identical positions we have identified to be invariant within prokaryotes and within eukaryotes but which differ between the two groups, four behave in still another manner in the mitochondrial initiator, while the three others fit the prokaryotic pattern. In particular, the semi-invariant position 11 (see Fig. 7 below) contains a pyrimidine in all known sequences except those of prokaryotic initiators where A is found; in the mitochondrial tRNA the base is U, while C is common to all eukaryotic initiators.

Therefore, if one accepts the proposal of Schwartz and Dayhoff (1978) that mitochondria and *Rhodospirillacae* share a recent common ancestor, it follows that mitochondrial tRNAs exhibit an unusually high mutation fixation rate. This rate may be reflected as well by the distant relationship betweeen the small subunit ribosomal RNA and prokaryotic 16S rRNA (Cunningham et al., 1977) and the inability of mitochondrial ribosomes to form hybrids with prokaryotic subunits (Griwell and Walg, 1972)

---

[2] The same is found for *N. crassa* mitochondrial $tRNA^{Tyr}$, whose sequence was recently published (Heckman et al., 1979).

even though both mitochondria and prokaryotes use formylmethionyl-tRNA protein synthesis initiation (Smith and Marcker, 1968).

On the other hand, the origin of the chloroplastic tRNA appears more straightforward. Chloroplastic phenylalanine tRNAs are clearly related to the blue-green algal sequence. Since Euglena cytoplasmic and chloroplastic tRNAs are most closely related to opposite branches (that is the prokaryotic and eukaryotic branches) of the $tRNA^{Phe}$ tree, direct evidence is presented supporting the endosymbiotic theory of the origin of chloroplasts contained in the eukaryotic hosts (see also Zablen et al., 1975; Bonen and Doolittle, 1976).

tRNA and the Random Hit Evolutionary Model

No time scale was assigned to the phylogenetic trees since it would be rather speculative to relate any node to the paleontological record. However, using equation:

$$2\lambda_b t = -3/4 \ln (1 - \frac{4}{3}\pi) \qquad (1)\ \text{(Nei,1975)}$$

where π is the proportion of different nucleotides between homologous genes, t their divergence time and λ_b a uniform probability of nucleotide substitution per site per year, it should be possible to estimate the total number of mutations accumulated since the eukaryote-prokaryote divergence. On the average, it is found:

$tRNA^{Phe}$: $\pi = 0.357$, from which $\lambda_b t = 0.242$

$tRNA^{fMet}$: $\pi = 0.319$, from which $\lambda_b t = 0.208$

Thus, a prokaryotic and a eukaryotic $tRNA^{Phe}$ should be separated from each other by about 37 mutations. The corresponding figure for $tRNA^{fMet}$ is about 31.5. However, a direct inspection of the two trees shows that the minimal number of mutations to be 10 to 20% higher than expected from Eq. 1. Moreover, the sparseness of these trees makes it likely that additional sequence would increase the number of detectable mutations. Finally, details of the mutation pattern (see below) make it clear that the random hit model used to derive Eq. 1 is not appropriate for tRNAs. Rather, the variance of λ is expected to be large since the various regions of the tRNA molecule differ sharply in their respective susceptibility to mutation; Eq. 1 will thus lead to a serious underestimate of the mutation rate, if distantly related sequences are compared. Here, branch lengths be directly used to compute a *minimal* nucleotide substitution rate; assuming the eukaryote-prokaryote divergence to be 1.8×10^9 years old (Kimura and Ohta, 1973), an *average* value $1.2 - 1.6 \times 10^{-10}$ mutations per site per year is found.

Since phylogenetic trees and derived ancestral sequences can be used to compute the minimum number of mutations that were fixed at any position, we have obtained these values for the two trees (see Fig. 3). Neither the $tRNA^{Phe}$ nor the $tRNA^{fMet}$ distribution is consistent with the single Poisson distribution resulting from a uniform probability of mutation at all positions. Rather, Fig. 3 reveals at least two types of position with different variabilities. From these distributions, it may be possible to find a best fit to a double Poisson distribution and obtain an estimate of the number of positions whose identity is an absolute requirement for biological function (Margoliash et al., 1971). An

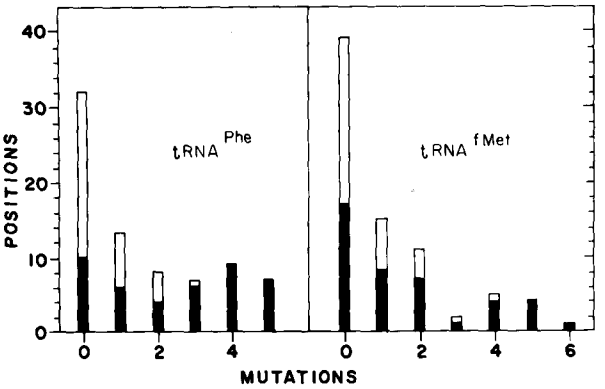

Fig. 3. Mutational 'spectra' of $tRNA^{Phe}$ and $tRNA^{fMet}$ show number of positions as a function of minimal number of nucleotide replacements per position. The shaded part indicates the base-paired regions in the standard cloverleaf structure, the white area, the rest of the molecule. Mitochondrial sequences are excluded.

accurate estimate of these true invariant positions is not however feasible in this case, since various models involving no constant residues at all could account for the actual distributions. Of course, such an assumption is also unrealistic since some residues must undoubtedly remain constant for functional reasons (CCA terminus, anticodon, and so on). Finally Fig. 3 reveals a clear distinction between regions which are base-paired in the cloverleaf structure and the rest of the molecule. The former contains a clear majority of the variable positions, a point whose functional implications are discussed below.

Types of Mutations

1° $tRNA^{Phe}$. Table 1 gives the distribution of nucleotide replacements classified according to type. Checking whether it agrees with the assumption of an equal probability for every possible nucleotide interchange shows this to be clearly not the case. The pattern is dominated by G → A, G → C and C → U changes. The transitions to transversions ratio is 0.89, significantly greater ($p \leqslant 0.002$) than the 0.5 expected from the equal probability hypothesis. Similarly, there are 82.5 replacements starting from G + C as opposed to 38.5 from A + U, significant at ($p \leqslant 0.001$). Substitutions converting G = C to A = U and C = G base pairs consitute 55% of the total against an expected one third. However, the apparent directionality of interchanges could simply reflect the high G + C content of the variable regions, mostly located in helical domains. To check this, we substracted the invariant and semi-invariant residues and recalculated the replacement pattern. With a calculated 61% G + C content in the variable regions of the ancestral sequence, a random mechanism predicts this same proportion of changes to start from these two nucleotides; actually, there are 76.5 changes starting from G + C against 36.5 from A + U (68%). Similarly, we find 50 replacements from G + C to A + U and 25.5 in the opposite direction; this imbalance can also be related to the high G + C content of the

Table 1. Nucleotide replacements in $tRNA^{Phe}$

From \ To	G	A	U	C
G	–	17.9 (14.4)	9.5 (9.5)	15.5 (15.5)
A	9.9 (8.4)	–	5 (5)	4.2 (4.2)
U	6 (6)	6 (6)	–	7.4 (6.9)
C	11 (11)	6.7 (6.7)	21.9 (19.4)	–

For each pair, the first figure gives the inferred number of changes when the whole sequence is considered; numbers in brackets refer to substitutions found in the variable part of the sequence when the so-called invariant and semi-invariant residues and the anticodon have been substrated. In standard nomenclature, these are positions 8.11,14,15,18,19,21,24,32 to 37, 48,53 to 58,60,61,74,75,76. (Rich and RajBhandary, 1976). Directionality of mutations results from assignment of the root of the tree to the eukaryote-prokaryote divergence.

Table 2. Nucleotide replacements in $tRNA^{fMet}$. Legend as in Table 1. Absence of a base is denoted by minus

From \ To	G	A	U	C	–
G	–	10 (9.5)	8.1 (7.6)	5.7 (5.7)	0
A	7.5 (6)	–	4.7 (3.7)	4.1 (3.6)	0
U	2.6 (1.5)	4.7 (3.7)	–	7.2 (6.2)	1.0 (1.0)
C	6.2 (6.2)	8.1 (7.6)	16.2 (15.2)	–	1.0 (1.0)
–	0	0	0.5 (0.5)	1.0 (1.0)	–

Total sample includes 40.9 transitions and 44.2 transversions. After subtraction of invariant and semi-invariant residues, G + C content of the ancestral sequence is 76.5%; this leaves 52.5 substitutions starting from G + C, of which 39.9 go to A + U, against a total of 25.9 starting from A + U, including 17.3 going to G + C.

ancestral sequence. Taken together, these results suggest that $tRNA^{Phe}$ still has not reached mutational equilibrium and undergoes a gradual replacement of G = C by A = U base pairs accompanied by a G = C to C = G interchange.

2° $tRNA^{fMet}$. Similar conclusions about $tRNA^{fMet}$ can be reached from inspection of Table 2. Lower cost of transitions over transversions and predominance of changes starting from G + C. Also, replacements from G + C to A + U occur half as often as in the opposite direction: The last two effects can again be explained as a result of a composition bias reflecting the high G + C content of the ancestral sequence. Although the individual types of interchanges are clearly not randomly distributed, only the C → U replacement, also the most common in $tRNA^{Phe}$ is clearly preponderant.

Finally the sum of Tables 1 and 2 produces a distribution quite similar to the one found earlier in the 5S RNA family (Sankoff et al., 1976), except for the much higher deletion-insertion cost in tRNAs.Combination of data in Table 3 gives a sizable sample

Table 3.

From \ To	G	A	U	C
G	–	46.9 (27.9)	31.1 (17.6)	31.0 (21.2)
A	32.4 (17.4)	–	18.0 (9.7)	18.1 (8.3)
U	15.1 (8.6)	16.0 (10.7)	–	32.9 (14.6)
C	26.0 (17.2)	24.1 (14.8)	67.9 (38.1)	–

Combined nucleotide replacement data from 5S RNA (Sankoff et al., 1976), $tRNA^{Phe}$ and $tRNA^{fMet}$. Insertions and deletions are not shown from the sample. The first figure summarizes all data, while the number in brackets refers to $tRNA^{Phe}$ plus $tRNA^{fMet}$ only. Total sample comprises:

- 180.1 transitions and 179.4 transversions
- 227 interchanges starting from G or C; of these, 170 go to A or U
- 132.5 interchanges starting from A or U, of which 98.5 go to G or C.
- C → U and G → A transitions as the most common types of nucleotide replacement.

whose characterisitcs could depend mostly on an interaction between the probability of each type of mutation and helix stability factors affecting its possible fixation or rejection. Also, this information could be used in future studies on similar types of structural RNAs to assign specific weights to every type of replacement, as is currently done in protein evolution studies (Dayhoff et al., 1972b).

Correlation with Secondary and Tertiary Structure

We have used the phylogenetic trees to derive a distribution of mutations on the secondary and tertiary structure map of these two tRNA families. Figure 4 shows such a simplified tertiary structure model of yeast $tRNA^{Phe}$ based on the X-ray structure (Kim, 1978). Immediately obvious is the high proportion of inferred mutations in the acceptor and TψC stem regions (nearly 2/3). Moreover, since reversions and parallel mutations are most likely to escape detection in 'hot spots', the real figure could easily be higher. Another variable region is found in the upper part of the anticodon stem. On the other hand, the D arm, the anticodon and TψC loops are highly conserved.

Here, the lower cost of transitions over transversions would support the idea that G = U and A = C pairs are frequently used as intermediates in base pairs interchanges. Indeed G = U pairs are quite common in tRNAs and sometimes persist in the course of evolution (Mizuno and Sundaralingam, 1978); A = C is much rarer but however do exist (Lomant and Fresco, 1975; Ninio, 1979b).

The conservation of the rest of the molecule is best explained by functional requirements. For example, TψC and anticodon loops are known to be involved in specific interactions with the ribosome and messenger RNA respectively. The three-dimensional structure of crystalline yeast $tRNA^{Phe}$ is best described as a right-angle intersection of two helical domains whose hinge is provided by additional tertiary interactions between the D and TψC loops each involving two invariant residues. This so-called 'augmented D helix', which involves both pairing and stacking interactions of the D arm with the

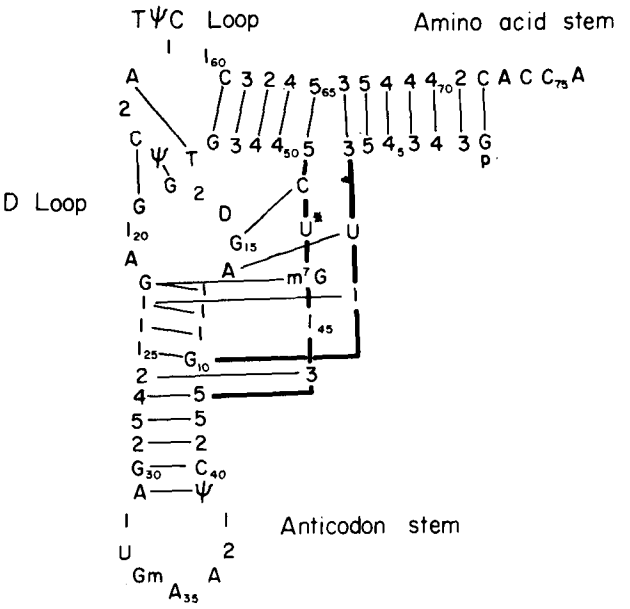

Fig. 4. Mutational map of $tRNA^{Phe}$. The molecule is drawn, according to Kim (1978), as a schematic model of yeast $tRNA^{Phe}$ tertiary structure in order to show hydrogen bonding, a large number indicates the number of mutations recorded at this position, a small number in subscript the position in the sequence. The parent base identifies invariant positions, except when its modified form predominates. Light lines indicate secondary and tertiary hydrogen bonding, heavy lines the sequences connecting the two helical domains

sequences connecting the acceptor to the D arm and the TψC to the anticodon arm, constitutes the brace that holds the structure together. (Rich and RajBhandary, 1976). Figure 4 shows most of these critical interactions to be preserved. Moreover, 7 out of the 8 mutations recorded in the D stem and in the 3 other positions forming the base triples occur in the *S. pombe* line alone, but molecular models show that the latter tRNA can still fit quite well into the standard conformation of $tRNA^{Phe}$.

The $m^7G_{46} = (G_{22} \equiv C_{13})$ 'triple' is replaced in *S. pombe* by $m^7G_{46} = (G_{22} = \psi_{13})$. However, the tertiary interaction of m^7G_{46} with G_{22} should not be strongly disturbed by the replacement of $(G_{22} \equiv C_{13})$ by a $(G_{22} = \psi_{13})$ 'wobble' pair (Fig. 5).

A G_9 -$(C_{23} \equiv G_{12})$ 'triple' in *S. pombe* substitutes for the usual $A_9 = (A_{23} = U_{12})$. The polynucleotide backbone of yeast $tRNA^{Phe}$ is quite extended at position 9 whose base is not sterically hindered to slide out (Goddard, 1977) and the slightly larger G_9 - $(C_{23} \equiv G_{12})$ 'triple' could be easily accomodated (Fig. 6) as already discussed for other tRNAs (Sigler, 1975; Brennan and Sundaralingam, 1976).

The eukaryotic G_{45} - $(m^2G_{10} \equiv C_{25})$ 'triple' (ψ_{25} in *S. pombe*) is impossible in prokaryotes where U_{45} replaces G_{45}.

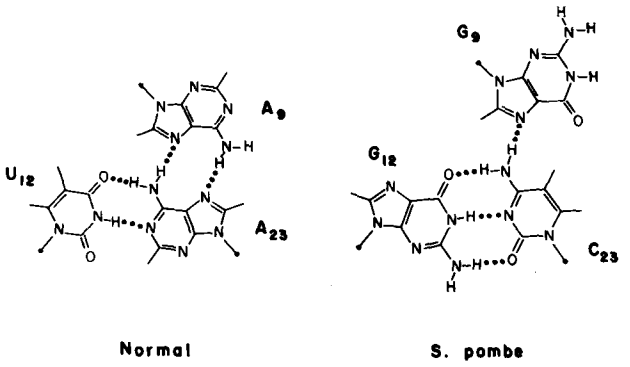

Fig. 5. Tertiary interaction in $tRNA^{Phe}$ sequences involving position 13, 22 and 46

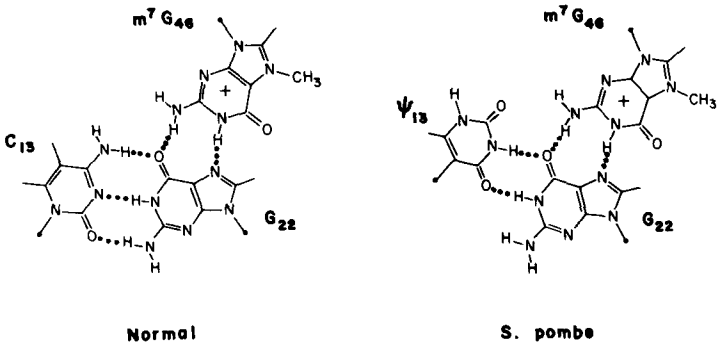

Fig. 6. Tertiary interaction in $tRNA^{Phe}$ sequences involving positions 9, 12 and 23

The mutation map of $tRNA^{fMet}$, as shown in Fig. 7, is strikingly similar – a high variability of the acceptor, TψC and upper anticodon stems is observed while most of the rest of the structure is preserved. The electron density map of crystalline yeast $tRNA^{fMet}$, which has been carried only to a resolution of 4.5 Å, nevertheless makes it clear that its sugar phosphate backbone architecture, except for the anticodom arm, is essentially identical to that of yeast $tRNA^{Phe}$ (Schewitz et al., 1979). Both molecules probably share similar tertiary interactions and, in this respect, it is not surprising that the 'augmented D helix' of $tRNA^{Met}$ undergoes only a very limited number of mutations. The tertiary pairs, excluding a $T_{54} = A_{58}$ to $A_{54} = m^1A_{58}$ change in eukaryotic $tRNA^{fMet}$, and the $m^7G_{46} = (G_{22} \equiv C_{13})$ 'triple' are preserved. It may also be noticed that 5'side of the D loop underwent 9 mutations, including deletions (or insertions). This is not surprising, since crystallographic data on yeast $tRNA^{Phe}$ have shown that outside

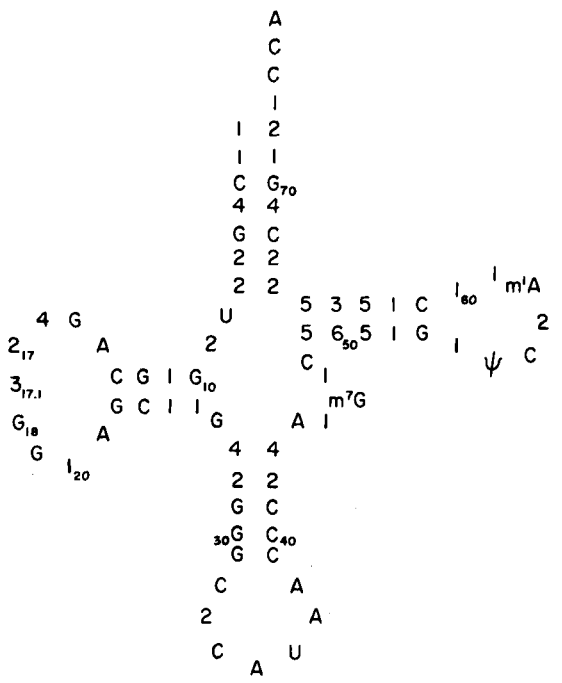

Fig. 7. Mutational map of $tRNA^{fMet}$, shown as a standard cloverleaf. Legend as in Fig. 4. Tertiary hydrogen bonding is omitted

of the molecule, a bulge is formed that could easily vary in length (Rich and RajBhandary, 1976). Finally, the lower part of the anticodon stem is constant in both $tRNA^{Phe}$ and $tRNA^{fMet}$ but sequence specific for each; the same two terminal base pairs are also found in mitochondrial sequences. Since this region is not formally involved in tertiary interactions, we hypothesize that it could constitute, along with the discriminator site at the fourth position from the 3' end (Crothers et al., 1972), part of the amino-acyl tRNA synthetase recognition site. This agrees with the general recognition of $tRNA^{Phe}$ and $tRNA^{Met}$ by the corresponding enzymes from heterologous systems and with most results which indicate the diagonal of the L shaped tRNA structure to interact with the synthetase (Rich and Schimmel, 1977). Although the acceptor stem is clearly involved in the case of *E. coli* $tRNA^{Tyr}$ (see review by Goddard, 1977), its variability probably excludes it from the recognition site for $tRNA^{Phe}$ and $tRNA^{fMet}$.

Our data thus lend further support to the current idea of a generalized conformation across all tRNAs consistent with their many common functions. The evolutionary process seems to have acted on initiator and phenylalanine tRNAs so as to preserve crucial conformational parameters while freely allowing replacements in most of the helical domains as long as arm-lengths and base-pairing are not modified.

Concluding Remarks

Transfer RNAs may represent one of the best subjects for the study of long term molecular evolution. Their ubiquity should allow derivation of a phylogenetic tree including all living species and ultimately provide a systematic approach for the study of genetic code evolution. Homology of all tRNAs is usually assumed from the high level of coincidence of all pairwise comparisons, but no rigorous proof of common descent has ever been put forward. Although homology may be shown in individual cases as in $tRNA^{Phe}$ and $tRNA^{fMet}$ families, a general proof may be beyond reach, since tertiary structure requirements constitute a strong barrier to changes in a large part of the molecule while other regions undergo repeated mutational events. Thus, sequence information can be used only with difficulty to relate different tRNAs, implying that random hit models are inappropriate to evaluate true divergence between sequence.

It follows that the application of a minimal mutational distance criterion to the construction of a tree including tRNAs with different functions may give misleading results since it is mostly 'noise' that will be observed. Here, we used the biological criteria, namely a unique acceptor or initiator activity and codon recognition, to determine which sequences should be compared. In doing so, a recently developed test of convergence applied to the two most distantly related groups of organisms, prokaryotes and eukaryotes, constitutes an internal check of homology which justifies this approach (Cedergren et al., 1979). At the present time, we have to conclude that there is no detectable homology left between mitochondrial and prokaryotic or eukaryotic sequences; the assumed universality of the genetic code, however, does not support the idea of an independent origin of mitochondria. Given this large divergence of the few known mitochondrial sequences, it is obvious that this problem could be solved only when enough tRNA data will allow reconstruction of a mitochondrial phylogenetic tree. It can also be stressed that much information could be gained from the study of so far neglected organisms: Anaerobic, non-O_2-evolving photosynthetic, and extreme halophilic bacteria. In addition, a better sampling of fungi and protozoa could add much to the present trees.

A particularily important finding of this work is the parallelism in the frequency of certain types of inferred mutations derived from these two tRNA families and the 5s family (Sankoff et al., 1976). Since tRNA and 5s RNA have vastly different functions implying different evolutionary constraints, that transitional mutations are more frequent than transversions must be the result of a fundamental mechanism of mutational fixation in RNA. We feel that this mechanism which applies particularily to base paired regions, the mutational 'hot spots' of tRNA and 5s RNA, can be explained by local steric and energetic (stacking) effects in helical regions of the molecule. Again this consideration implies that 'mismatches' due to mutation are generally of the type G = U or A = C rather than pyrimidine-pyrimidine or purine-purine (Mizuno and Sundaralingam, 1978; Lomant and Fresco, 1975; Ninio, 1979b).

As a final remark, we have as yet not taken into account the presence and the identity of modified nucleosides in our analysis. Indeed, it has been proposed that nucleoside modification variation may be more sensitive to phylogenetic relationship than the sequence (Cedergren and Cordeau, 1973; Ninio, 1979a). We are currently evaluating this hypothesis which could add valuable information in the derivation of tRNA evolu-

tionary relationship. Likewise, we have not evaluated the possible effect of intervening sequences in certain eukaryotic tRNA genes (Knapp et al., 1978; Valenzuella et al., 1978; O'Farrell et al., 1978), although the results presented here dealing primarily with the functional molecule should not be affected seriously by the existence of non-continuous genes.

Acknowledgments. We thank Drs. Clarkson, Dirheimer, Olins, Rabinowitz, Delk, Oshima and RajBhandary for communicating to us their sequences prior to publication. We also thank the Fonds National de la Recherche Scientifique Belge for a grant to support a short term stay of Henri Grosjean in Montréal.

References

Bonen, L., Doolittle, W.F. (1976). Nature **261**, 669–673

Brennan, T., Sundaralingam, M. (1976). Nucl. Ac. Res. **3**, 3235–3251

Cedergren, R.J., Cordeau, J.R. (1973). J. Theoret. Biol. **39**, 477–486

Cedergren, R.J., Sankoff, D., LaRue, B., Grosjean, H. (1979). Proc. Natl. Acad. Sci. USA (in press)

Crothers, D.M., Seno, T., Söll, D.G. (1972). Proc. Nat. Acad. Sci. USA **69**, 3063–3067

Cunningham, R.S., Gray, M.W., Doolittle, W.F., Bonen, L. (1977). In: Acides Nucléiques et Synthèse des Protéines chez les Plantes, p. 7. Colloques internationaux C.N.R.S., no 261. Paris: Editions du Centre National de la Recherche Scientifique

Dayhoff, M.O., Park, C.M., McLaughlin, P.J. (1972a). In: Atlas of protein sequence and structure, vol. 5, M.O. Dayhoff, ed., p. 7 Silver Spring, Maryland: National Biomedical Research Foundation

Dayhoff, M.O., Eck, P.V., Park, C.M. (1972b). In: Atlas of protein sequence and structure, vol. 5, M.O. Dayhoff, ed., p. 89. Silver Spring, Maryland: National Biomedical Research Foundation

Dirheimer, G., Keith, G., Sibler, A., Martin, R. (1979). In: Cold Spring Harbor Monograph **9A**: tRNA, structure, properties and recognition. Schimmel, P., Söll, D., Abelson, J. (eds.) (in press)

Gauss, D., Grüter, F., Sprinzl, M. (1979). Nucl. Ac. Res. **6**, r1-r9

Goddard, J-P. (1977). Prog. Biophys. Mol. Biol. **32**, 233–308

Grivell, L.A., Walg, H.L. (1972). Biochem. Biophys. Res. Commun. **49**, 1452–1458

Hasegawa, M. (1978). Orig. Life **9**, 495–500

Heckman, J.E., Hecker, L.I., Schwartyback, S.D., Barnett, W.E., Baumstark, B., RajBhandary, U.L. (1978). Cell **13**, 83–95

Heckman, J.E., Alzner-Deweerd, B., RajBandary, U.T. (1979). Proc. Natl. Acad. Sci. USA **76**, 717–721

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., Zamir, A. (1965). Science **147**, 1462–1465

Holmquist, R., Jukes, T.H., Pangburn, S. (1973). J. Mol. Biol. **78**, 91–116

Kim, S.H. (1978). In: Transfer RNA, S. Altman ed., p. 248. Cambridge, Mass.: The MIT Press

Kimura, M., Ohto, T. (1973). Nature New Biol. **243**, 199–200

Knapp, G., Beckmann, J., Johnson, P., Fuhrman, S., Abelson, J., (1978). Cell, **14**, 221–236

Lomant, A.J., Fresco, J.R. (1975). In: Progr. Nucl. Ac. Res. and Mol. Biol., vol. 15, p. 185–218, Cohn, W.A. ed. New York: Academic Press
Margoliash, E., Fitch, W.M., Dickerson, R.E. (1971). In: Biochemical evolution and the origin of life, E. Schoffeniels, ed., p. 52. Amsterdam: North-Holland
Martin, R., Sibler, A., Sneller, G., Keith, G., Stohl, A., Dirheimer, G. (1978). Nucl. Ac. Res. 5, 4579–4592
Mizuno, H., Sundaralingam, M. (1978). Nucl. Ac. Res. 5, 4451–4461
Moore, G.W., Goodman, M., Barnabas, J. (1973). J. Theroret. Biol. 38, 423–457
Nei, M. (1975). In: Molecular population genetics and evolution, A. Neuberger and E.L. Tatum, eds; p. 62–82, Amsterdam: North-Holland
Ninio, J. (1979a). In: Approches Moleculaires de l'Evolution, chap. III and XV. Paris: Masson
Ninio, J. (1979b). Biochimie, (in press)
O'Farrell, P., Cordell, B., Valenzuella, P., Rutter, W., Goodman, H. (1978). Nature 274, 438–440
Rich, A., RajBhandary, U.L. (1976). Ann. Rev. Biochem. 45, 805–860
Rich, A., Schimmel, P. (1977). Nucl. Ac. Res. 4, 1649–1665
Sankoff, D., Cedergren, R.J., Lapalme, G. (1976). J. Mol. Evol. 7, 133–149
Schewitz, R.W., Podjarny, A.D., Krishnamachari, N., Hughes, J.J., Sigler, P.B., Sussman, J.L. (1979). Nature 278, 188–190
Schwartz, R.M., Gantl, M.J., Dayhoff, M.O. (1976). In: Atlas of protein sequence and structure, suppl. 5 to vol. 2, M.O. Dayhoff, ed., p. 271. Silver Spring, Maryland: National Biomedical Research Foundation
Schwartz, R.M., Dayhoff, M.O. (1978). Science 199, 395–403
Sigler, P. (1975). Ann. Rev. Biophys. Bioeng. 4, 477–527
Silverman, S., Heckman, J., Cowking, J., Delaney, A., Dunn, R., Gilman, J., Tener, G., Söll, D., RajBhandary, U.L. (1979). Nucl. Ac. Res. 6, 421–433
Smith, A.E., Marcker, K.A. (1968). J. Mol. Biol. 38, 241–248
Valenzuella, P., Venegas, A., Weinberg, F., Bishop, R., Rutter, W. (1978). Proc. Natl. Acad. Sci. USA 75, 190–194
Zablen, L.B., Kissil, M.S., Woese, C.R., Buetow, D.E. (1975). Proc. Nat. Acad. Sci. USA 72, 2418–2422

Received August 5, 1979

Note Added in Proof

The lower part of the $tRNA^{fMet}$ anticodon stem cannot be a specific synthetase recognition site, since the initiator and the non-initiator methionine tRNAs are usually both recognized by the same enzyme but have different sequences in this region. Rather, recent results seem to indicate that the two constant G ≡ C pairs help to maintain the unique confirmation found in the anticodon loop of initiator tRNAs (Wrede, P., Woo, N.H., Rich, A. (1979). Proc. Natl. Acad. Sci. USA 76, 3289–3293