THE EVOLVING tRNA MOLECULE

Authors: **R. J. Cedergren**
Department of Biochemistry
David Sankoff
Center for Mathematical Research
University of Montreal
Quebec, Canada

Bernard LaRue
Department of Chemistry and Biology
University of Quebec
Trois-Rivieres
Quebec, Canada

Henri Grosjean
Laboratory of Chemical Biology
Free University of Brussels
Brussels, Belgium

Referee: Dieter Söll
Department of Biophysics and Biochemistry
Yale University
New Haven, Connecticut

I. AN INVITATION TO tRNA EVOLUTION

". . . il n'est pas de vraie création sans secret." Albert Camus, *Le mythe de Sisyphe*

Interest in tRNA biochemistry is attested to by the imposing mass of data generated by a veritable army of researchers. The approach to this information used here will be somewhat unique among the large number of existing tRNA reviews in that it will be organized and interpreted within an evolutionary framework. In addition to a discussion and an extension of current hypotheses on tRNA evolution, an evaluation is presented of the basic premise of this review: that tRNA structural and functional data can provide information on early evolutionary events.

The importance of tRNA evolution lies in its relation to the origin and development of the genetic code.[1] In a sense, the discovery of the universality of the genetic code in eukaryotes and prokaryotes (see Section VII. B. on mitochondria) was a mixed blessing to the field of molecular evolution.[2] On one hand it clarified the basis of protein sequence comparisons involving different organisms, but on the other it limited the applicability of comparative methods for identifying primitive vs. advanced forms. Thus, genetic code evolutionary models have had to be based on indirect evidence and/or speculation. The primary structure of tRNA, however, does vary, not only as a function of the particular amino acid and codon it relates, but also from species to species across the phylogenetic spectrum.[3] The authors hope to show in this review how statistical and comparative techniques in general can help to shed light on both Darwinian and pre-Darwinian evolution, i.e., biological evolution, speciation, and the biochemical evolution which preceded it.

Along with a comparative study of tRNA sequences three other interrelated aspects of

tRNA evolution will be stressed: the interactions of tRNA with other coevolving components of the machinery of protein synthesis, tRNA conformation, and the dynamic nature of tRNA function. Although tRNA may be thought of as the pivotal molecule in the fundamental process of protein synthesis, its structure, function and evolution must be considered in terms of the entire process, since it is the process which is subject to the forces of biological selection.

A. Molecular Evolution

In an era of extreme specialization in biology, the theory of evolution is unique in providing a basis for the integration, unification, and comprehension of various life forms in terms of generalized concepts.[4] In particular, the study of molecular evolution is a framework in which various molecular — biochemical or genetic — data of very diverse origins may be analyzed to derive widely valid principles. Although the idea of studying evolution at the molecular level had previously been proposed by Zuckerkandl and Pauling,[5] it was through the discoveries of macromolecular biology in the early 1960s that the mechanistic basis of molecular change was understood, providing the theoretical foundations for the field of molecular evolution. Thus the concept of a slow, continuous change in organismal characteristics could be understood in terms of cumulative, incremental changes on the molecular level.[6]

1. Dogma-DNA-Darwin

With the discovery of the molecular events involved in the expression of genes and genetic information flow, molecular biology was endowed with a dogma almost before it was recognized as a separate discipline: the genome of a cell, its DNA, stores genetic information which, after replication, can be distributed to its offspring. To express genetic information, messenger RNA is faithfully transcribed from DNA as a first step. Finally, the message encoded in the mRNA, in conjunction with the ribosome, transfer RNA, amino acids, and many other factors, is translated into protein. These steps hold for all organisms except certain RNA viruses that either replicate at the RNA level, thereby avoiding DNA completely, or can be integrated into the cell genome by first directing the transcription of DNA from their RNA.[7,8]

The entire process of genetic expression is relatively error-free due to the fidelity of the components and, in the case of DNA replication, to enzymatic error-correcting mechanisms.[9] Despite this stability, permanent structural changes in DNA do occur through the process of mutation whereby errors are incorporated into DNA during replication and/or repair.[4] The direction of information flow is such that only errors in DNA, provoked by the mutational process, permit variability in the germ line. This process can give rise to speciation through the physical isolation of subpopulations followed by distinct mutational events in the different subpopulations.[6] This divergent evolutionary model leads to the generally accepted idea that at some point in the past, probably more than 3.5 billion years ago, there existed a protoorganism or a colony of protoorganisms whose genes evolved by mutational processes to give rise to most present genes.[1]

One might ask whether all natural mutations are incorporated into DNA. Here it is thought that many are called, but few are chosen (fixed in the germ line). At least some mutations are thought to be fixed based on the selective advantage of the structural change (neo-Darwinian).[6] Unfixed mutations are thought to be deleterious. A considerable controversy has arisen, however, concerning the selective nature of fixed mutations. A large amount of data on protein evolution can be interpreted in terms of a neutral, non-Darwinian, mutational process where the incorporation of most mutations would have little or no selective value.[10] Although various mathematical models have been elaborated, there is not yet compelling evidence for a predominantly selectionist or

neutralist mechanism.[11] It is clear, moreover, that selectivity of a mutation must be considered at every step during gene expression and not only with regards to the biological activity of the gene product.

2. Evolutionary Inference Based on Proteins

Since protein structure is related to genome structure, the comparison of protein sequences having the same activity reflects to some extent phylogenetic relationships between different species. It will be valuable for us to underline some of the major findings in protein evolution. Among the most studied proteins, the cytochrome *c* molecule was the basis for many of the early papers. Statistical tests developed largely by Margoliash and colleagues showed that the various cytochrome *c* molecules are homologous, since their sequences are more related than they would have to be for reasons of biological activity.[12] Homologous* in this sense indicates that the protein set has diverged from a common ancestor. This justified placing these molecules into an evolutionary tree diagram to indicate their relatedness. The techniques for constructing such trees were developed by Fitch and Margoliash,[13] Dayhoff and McLaughlin,[1] and the school of numerical taxonomists (cf. Sneath and Sokal[14]). A striking result is that the protein taxonomy thus obtained was largely consistent with phylogenetic trees based on other criteria.[11] Protein taxonomy could, therefore, be used as an alternate method for establishing the evolutionary phylogeny of various organisms. This approach would be particularly useful for sets of organisms which are difficult to place in standard phylogenies because of lack of fossil data or other criteria for classification. It has been pointed out that although data on morphology and other properties of organisms have long been used to obtain reasonable phylogenies, there is a much greater danger, compared to adequate molecular data, of error due to convergent evolution and similar phenomena.[11]

It should be noted that gene duplication can complicate the evolutionary analysis. Thus the myoglobin-hemoglobin protein family is the result of an ancient gene duplication giving rise to two coexisting parallel branches (myoglobin and hemoglobin branches) which were subjected to differential evolutionary pressures.[1] The relationship between these two protein subfamilies has been termed paralogous, whereas the relationship between members of a particular branch was called orthologous.[12]

3. Evolving Nucleic Acids

The study of molecular evolution of nucleic acids has advantages and disadvantages compared to the study of protein evolution. As mentioned above, only DNA structural change can be established in a population, so the evolution of other macromolecules, RNA and proteins, is only pertinent to cellular evolution in that it can be related to DNA structure. DNA structure itself has been, until recently, difficult to determine but, now that improved methods are available great interest in DNA evolution can be expected.[15] Protein structure, on the other hand, is related to DNA only through the genetic code. Since some amino acids are coded by more than one codon, silent mutations can occur. Some amino acid substitution can also be explained by more than one set of mutational events at the DNA level. RNA structure, on the other hand, either messenger, ribosomal, or transfer, is directly related to DNA structure by the Watson-Crick rules of base-pairing.

The molecular events in RNA evolution can be difficult to infer, however, since an insertion or a deletion of nucleotides occasionally can occur during the history of a particular gene. It then becomes a problem to align correctly homologous regions of

* The term homology can be used only when structures share a common ancestor. Similarity per se is not proof of homology, since similar structures could arise by convergent evolution.

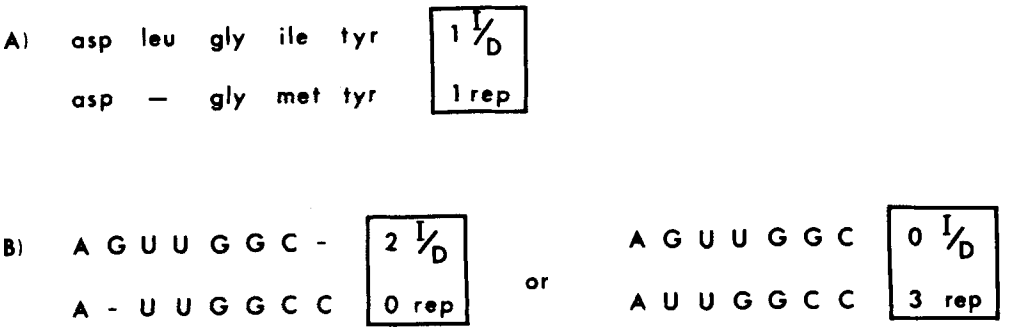


FIGURE 1. Aligning sequences with insertions or deletions. (A) Protein sequences made up of 20 possible elements can be aligned with relative ease. (B) The hypothetical nucleic acid sequence shown here could have two patterns of alignment; one which involves three replacements (rep), but no insertions or deletions (I/D). In order to properly evaluate which alignment in (B) represents the true evolutionary relationship, the alignment procedure could take into account the relative probability of replacements to insertions/deletions or be related to random sequence simulations.[16]

sequences used for determining phylogeny.[16] In protein sequences using a 20-letter alphabet, it is relatively easy to decide where and how large an insertion or a deletion should be whereas in nucleic acids, based on only a 4-letter alphabet, the alignment problem is much more complicated (see Fig. 1 and Section II.B.). In tRNA structure, however, the presence of constant nucleosides in known positions and the existence of the cloverleaf base-pairing arrangement permit a relatively easy alignment of these sequences.[17]

4. Functions of tRNA

The principal biological activity of tRNA is to transport amino acids to the ribosome, the site of protein synthesis. The amino acids are covalently attached to the terminal adenosine of the tRNA by means of an ester linkage. This reaction is catalyzed by aminoacyl-tRNA synthetases (ARS) or ligases. There is usually one of these in each cell for each of the 20 amino acids used in the genetic code, while the number of tRNAs specific for a particular amino acid is variable.[18]

Extrapolating from its crucial function during protein synthesis, tRNA and a close relative of the present-day genetic code would have to have been functional in the first modern-type cell or population.[1] Phylogenetically then, one should be able, at least in principle, to relate structures in present-day organisms to that of the ancestral organism.

Furthermore, it can be hoped that tRNA phylogenetic evidence may permit a direct approach to the evolution of the genetic code since, as the code evolved through increasingly refined versions, appropriate tRNAs would have had to be available to translate it and tRNA evolution should mimic and reflect that evolution. Although the study of tRNA evolution has the potential of responding to some of the above hopes, one may well ask whether enough relics of the past can now be identified within present-day structures to be able to reconstruct molecular events leading to the current genetic code or whether the pessimistic view of Holmquist et al.[19] that early evolutionary events are hidden by a long history of superimposed, random mutations, is justified.

To speak about tRNA evolution in solely molecular terms does, however, sidestep the question of real biological significance. tRNA does not evolve in a vacuum; it evolves as a response to, or rather it is selected according to its environment. While the evolution of the use of gears in a machine can be chronicled, their function and interaction with the other parts of the machine must not be ignored. The tRNA situation is much more complex than the case of protein evolution, since tRNA is called upon to serve many roles

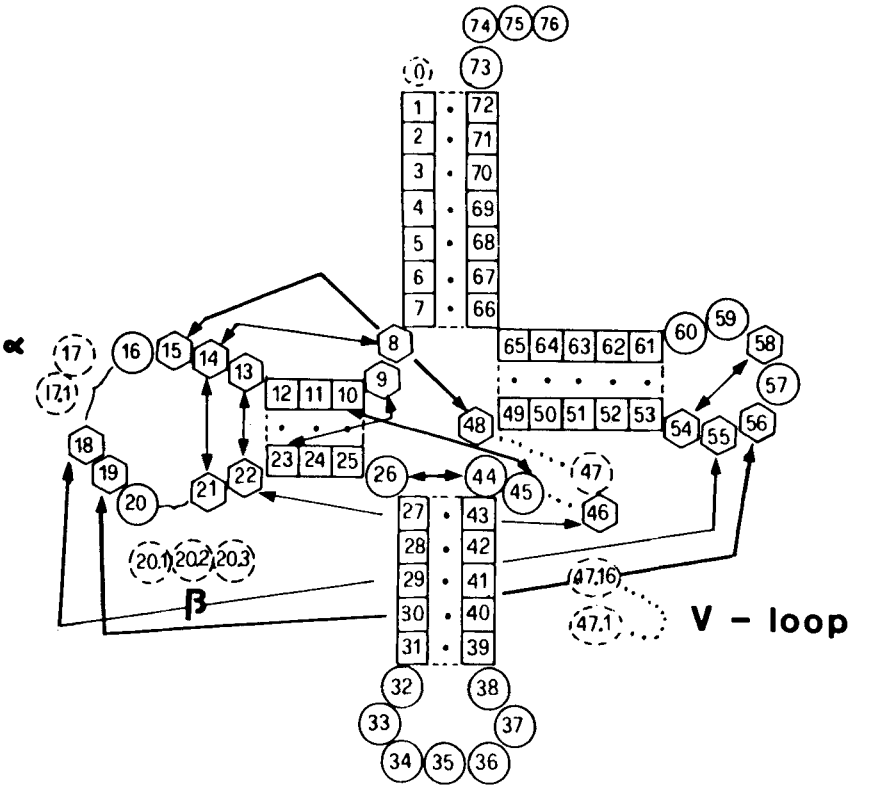


FIGURE 2. Generalized cloverleaf conformation of tRNA. Positions in the cloverleaf are shown using the numbering system adopted at the 1978 Cold Spring Harbor, New York, meeting on tRNA. Positions involved in base pairing are indicated by boxes with dots between complementary nucleotides. Hexagons indicate positions involved in tertiary structural interactions of yeast $tRNA^{Phe}$; arrows indicate the positions participating in such interactions. Circles are used to indicate unpaired positions not involved in base-base tertiary interactions. Dotted circles are used for positions which are not filled in all sequences.

in the modern cell. It interacts with a multitude of different molecules. In protein synthesis alone, tRNA is required to react with an amino acid and interact with an aminoacyl-tRNA synthetase, messenger RNA, the ribosome, plus a large number of protein synthesizing factors.[20] But tRNAs perform in other processes including cell wall synthesis,[21] nonribosomal aminoacylation of proteins,[22] acting as corepressors,[23] priming reverse transcriptase,[24] and so on. In addition, considering only the biosynthesis of tRNAs, correct interactions must be made with endo- and exonucleases, some 60 nucleoside-modifying enzymes, etc. One must, therefore, speak of an interactive tRNA coevolving with its environment.

5. *A Summary of tRNA Structure*

For further information on tRNA structure the reader should refer to the reviews published in recent years by various authors.[3,20,25-30] In summary, all tRNAs are made of a single polynucleotide chain, 73 to 93 residues in length, which can fold back upon itself to form the hydrogen-bonded cloverleaf structure initially proposed by Holley et al. for $tRNA^{Ala}_{IGC}$*.[31] This constant folding pattern sets the basis for a common system of positional alignment and numbering. Figure 2 describes the nomenclature which was

* Throughout this review the abbreviation $tRNA^{aa}_{xxx}$ is used to represent the tRNA specific for the amino acid aa, and having the anticodon sequence xxx.

adopted at the Cold Spring Harbor meeting on tRNA in 1978 and will be used throughout this article. All sequences of tRNA used for this review are shown in Table I (cf. Reference 3).

The cloverleaf conformation results in a natural division of the tRNA molecule into the five regions listed below, two of which, the acceptor and anticodon arms, are involved in its adaptor function:

1. The acceptor arm, which contains both the 3′ and 5′ ends, is a seven base-pair stem supporting the unpaired $XCCA_{OH}$ terminus. Upon tRNA amino-acylation, the amino acid is esterified to the 2′- or 3′-hydroxyl group of the 3′ terminal adenosine.[32]
2. The D arm is a three or four base-paired stem completed by a loop of variable length containing five constant residues. The two variable regions, α and β, are pyrimidine-rich and usually contain the reduced form of uridine, dihydrouridine.
3. The anticodon arm contains, at the end of a five base-paired stalk, a loop of seven residues whose third, fourth, and fifth positions constitute the anticodon. This binds to its complementary messenger RNA code triplet in an antiparallel fashion. The anticodon is flanked by two pyrimidines on its 5′ side and by a purine, frequently a hypermodified adenosine derivative, on its 3′ side. Residue 33 is always U except in some eukaryotic initiators where it is replaced by C. Position 34 is called the "wobble" position of the anticodon due to the occasional nonstandard base-pairing interactions possible with this residue during codon reading.[33]
4. The extra arm, 3 to 21 residues in length, plays an important role in helping to stabilize the tertiary structure.
5. The T arm is named after the quasi-invariant TΨC sequence found on the 5′ side of its seven residue loop. Its five base-pair stem always terminates with a G_{53}-C_{61} pair. The TΨC-purine sequence is possibly involved in tRNA interaction with prokaryotic 5S[34,35] or eukaryotic 5 8S[36] RNA on the large ribosomal subunit.

Some tRNAs may depart from the standard structure described above:

1. The 5′ terminal residue of the acceptor arm is unpaired in in vivo formylated prokaryotic initiator tRNAs, a feature responsible for their lack of binding to elongation factor Tu.[37] An exception to this initiator rule is the initiator $tRNA^{Met}$ from *Halobacterium cultirubrum,* which is paired and nonformylated in vivo.[20] Another departure from the generalized structure involves $tRNA^{His}$ from *Escherichia coli* or *Salmonella typhimurium* which has an extra 5′ nucleotide paired to the fourth position from the 3′ end, giving eight base pairs in the acceptor stem.[38]
2. In yeast mitochondrial $tRNA^{Phe}$, the T stem contains an extra residue which has to bulge out if a normal five base-pair helix is to be formed.[39]
3. The UGA suppressor $tRNA^{Leu}$ from *Schizosaccharomyces pombe* has an anticodon loop of nine nucleotides owing to an unpaired A-C at the first position of the supporting stem;[40] this unusual feature is also found in some *E. coli* su_2^+ strains having a mutant $tRNA^{Gln}$.[41] The codon-anticodon interaction, however, seems to take place as usual since these latter tRNAs act as normal nonsense suppressors, suggesting that their anticodon loop conformation is not drastically modified. In contrast, the insertion mutant $tRNA^{Gly}$ from *Salmonella typhimurium* has an eight nucleotide anticodon loop and behaves as a frameshift suppressor in reading the quadruplet GGGG as a codeword.[42]
4. The first and last base of the T loop in $tRNA_1^{Val}$ from rabbit liver have been shown to be paired in an isolated 3′ terminal fragment of the tRNA.[43] However, since T_{54} and m^1A_{58} of yeast $tRNA^{Phe}$ are involved together in a crucial tertiary interaction which

Table 1
KNOWN tRNA SEQUENCES

Amino acid	Codons (5′—3′)	Anticodons (5′—3′) — species		
		Prok.	Euk.	Others
ILE	AUU	GAU ECO	IAU TUT	UAU T_4
	AUC			
	AUA			
LEU	CUU	GAG ECO	UAG SCE	UAG T_4
	CUC	CAG ECO	CAG XEN	
	CUA			
	CUG			
	UUA	AAA ECO	UAA SPO	
	UUG		CAA SCE	
			CAA TUT	
LYS	AAA	UUU ECO	UUU SCE	
	AAG	UUU BSU	CUU SCE	
			UUU MAM	
			CUU MAM	
			CUU MAM	
			CUU DRO	
MET-M	AUG	CAU ECO	CAU SCE	
		CAU BSU	CAU MAM	
MET-F	AUG	CAU ECO	CAU SCE	CAU MT-NCR
		CAU BSU	CAU NCR	CAU MT-SCE
		CAU TTH	CAU WGE	
		CAU TTH	CAU VER	
		CAU MYC	CAU DRO	
		CAU SFE	CAU SOB	
		CAU ANI		
PHE	UUU	GAA ECO	GAA SCE	GAA CH-EGR
	UUC	GAA BSU	GAA WGE	GAA CH-BEA
		GAA BST	GAA EGR	GAA MT-SCE
		GAA MYC	GAA MAM	
		GAA ANI	GAA SPO	
			GAA DRO	
ALA	GCU	UGC ECO	UGC SCE	UGC MT-NCR
	GCC	UGC BSU	IGC TUT	
	GCA	GGC ECO	IGC BMO	
			UGC XEN	
ARG	CGU	ICG ECO	ICG SCE	
	CGC		ICG MAM	
	CGA			
	CGG			
	AGA		UCU SCE	UCU T_4
	AGG			
ASN	AAU	GUU ECO	GUU MAM	
	AAC			
ASP	GAU	GUC ECO	GUC SCE	
	GAC			
CYS	UGU	GCA ECO	GCA SCE	
	UGC			
GLN	CAA	UUG ECO		UUG T_4
	CAG	CUG ECO		
GLU	GAA	UUC ECO	UUC SCE	
	GAG		UUC SPO	
GLY	GGU	GCC ECO	GCC SCE	UCC T_4
	GGC	UCC ECO	GCC WGE	UCC[a] SEP
	GGA	CCC ECO	GCC BMO	UCC[a] SEP

Table 1 (continued)
KNOWN tRNA SEQUENCES

Amino acid	Codons (5′—3′)	Anticodons (5′—3′) — species		
		Prok.	Euk.	Others
	GGG	GCC BSU	UCC BMO	
		UCC BSU	GCC HUM	
			CCC HUM	
HIS	CAU	GUG ECO		GUG MT-SCE
	CAC			
PRO	CCU		IGG MAM	UGG T_4
	CCC		UGG MAM	
	CCA			
	CCG			
SER	UCU	UGA ECO	IGA SCE	UGA T_4
	UCC		IGA SCE	UGA MT-SCE
	UCA		UGA SCE	
	GCG		CGA SCE	
			UGA SPO	
			IGA RAT	
	AGU	GCU ECO	GCU RAT	
THR	ACU	GGU ECO	IGU SCE	UGU T_4
	ACC	UGU BSU	UGU MAM	UAG[b] MT-SCE
	ACA			
	ACG			
TRP	UGG	CCA ECO	CCA SCE	
			CCA MAM	
			CCA CHI	
TYR	UAU	GUA ECO	GUA SCE	GUA MT-NCR
	UAC	GUA BST	GUA TUT	
		GUA BSU	GUA SPO	
			GUA XEN	
VAL	GUU	GAC ECO	IAC SCE	UAC MT-SCE
	GUC	GAC ECO	UAC SCE	
	GUA	UAC ECO	CAC SCE	
	GUG	GAC BST	IAC TUT	
		UAC BSU	IAC MAM	
			CAC MAM	

Note: ECO = *Escherichia coli;* BSU = *Bacillus subtilis;* BST = *Bacillus stearothermophilus;* TTH = *Thermus thermophilus;* MYC = *Mycoplasma;* SFE = *Streptococcus faecalis;* ANI = *Anacystis nidulans;* SCE = *Saccharomyces cerevisiae;* TUT = *T. utilis,* T_4-Phage T_4; BMO = *Bombyx mori;* XEN = *Xenopus laevis;* MAM = mammal; SPO = *Schizosaccharomyces pombe;* WGE = wheat germ; DRO = *Drosophila;* NCR = *Neurospora crassa;* VER = vertebrate; SOB = *Scenedesmus obliquus;* CHI = chicken; MT = mitochondrial; CH = chloroplastic.

[a] Does not participate in protein synthesis.
[b] Recognizes a leucine codon.

From Sprinzl, M., Grueter, F., Spelzhaus, A., and Gaus, D.H., *Nucl. Ac. Res.,* 8, r1, 1980. With permission.

stabilizes interactions between the D and T loops,[20] the significance of this observation with regard to the intact $tRNA_1^{Val}$ structure is unclear.

5. In addition to the standard Watson-Crick pairs, many tRNAs contain in their stems a few, usually only one per molecule, nonstandard pairs of which G-U is the most frequently found.[29] Figure 3 shows the identity and location of these pairs. The

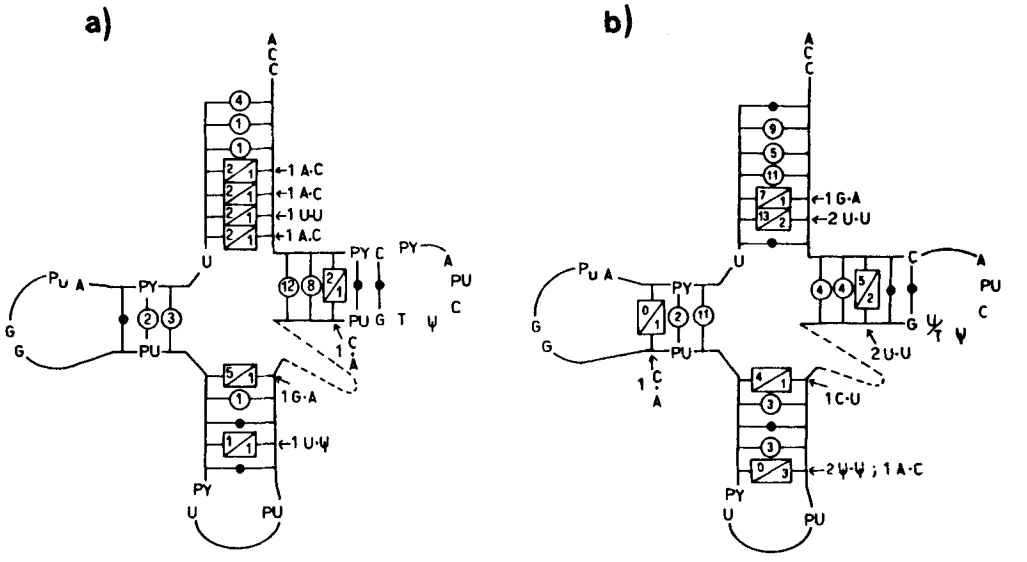


FIGURE 3. Unusual base pairs in tRNA. The location of unusual base pairs in the cloverleaf conformation of tRNA is shown. Invariable and semi-invariable positions are identified as positions where only purines (pu) or pyrimidines (py) are permitted. Encircled numbers indicate the frequency of G-U pairs. Boxes between nucleosides indicate that other than G-U pairs are permitted. The frequency of G-U pairs in these positions is shown in the upper left of the box, whereas the lower right indicates the occurrence of other unusual pairs. The unusual pair is noted next to position involved. Dots are placed between nucleosides in positions that are invariably Watson-Crick pairs. (a) Composite of 42 known prokaryotic tRNA sequences plus 8 T_4-coded tRNAS; (b) composite of 62 eukaryotic sequences.

helical conformation of the stems are probably not significantly disturbed by the occurrence of a single unusual pair.

6. Quite a number of tRNAs, especially mitochondrial, differ from the standard structure in that one or more of the "constant" nucleotides are missing. These exceptions have been reviewed in detail elsewhere,[39] and some of these will be examined here when they can be related to evolutionary processes.

Heterogeneity in tRNA length arises from the variable number of bases in the D loop and in the extra arm (Figure 2), but this variability is not thought to involve radical alterations in tRNA conformation. The pairing scheme of the cloverleaf and the presence of invariant and semi-invariant nucleotides gave evidence for the idea of a common tertiary structure for all tRNAs. That prediction was largely confirmed by X-ray diffraction data from yeast $tRNA^{Phe\,30}$ to which can be added the recently determined structures of yeast[44] and *E. coli* initiator tRNAs.[45] The tertiary structure of yeast $tRNA^{Phe}$, which is so far the most detailed, is illustrated by the diagrams of Figure 4. Its most pertinent features are summarized as follows:[20,27-30]

1. The molecule has the general shape of an L and incorporates all of the secondary hydrogen bonding predicted by the cloverleaf conformation.
2. The acceptor and T stems are stacked on each other and form a continuous double helix. An analogous arrangement is also found with the anticodon and D stems which sandwich a nonstandard $m_2^2G_{26}$-A_{44}* tertiary pair.

* m_2^2G represents 2-dimethylguanosine. For all symbols identifying modified nucleosides see Section V. Throughout the review the convention that a dash between bases indicates base pairing and a dot indicates tertiary interactions is used.

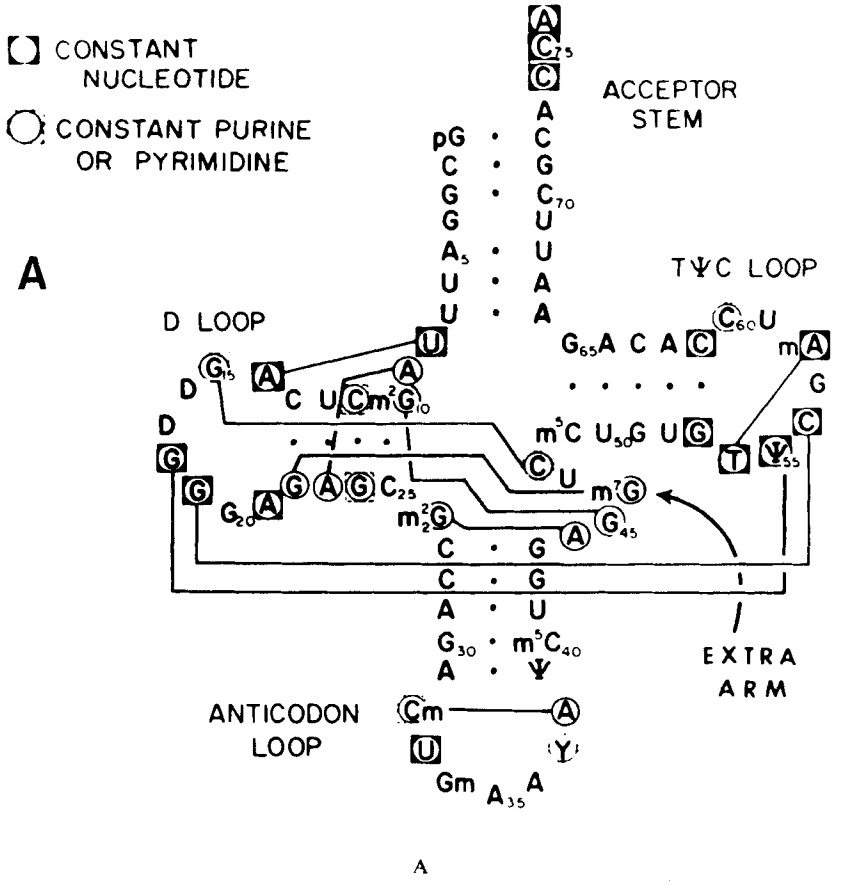


FIGURE 4. Tertiary structure of yeast $tRNA^{Phe}$. (A) The sequence of yeast tRNA in the cloverleaf conformation is shown. Invariable and semi-invariable positions are indicated by full and dotted boxes, respectively. Tertiary structure interactions are indicated by heavy lines joining the nucleosides involved. (From Rich, A. and RajBhandry, U. L., *Annu. Rev. Biochem.*, 45, 805, 1976. With permission). (B) The ribose-phosphate backbone in the crystal conformation is schematized by the continuous tube. Tertiary base pairs are indicated by black rods; other rods show the base pairs found in the standard cloverleaf conformation. (From Kim, S. H., *Transfer RNA*, MIT Press, Cambridge, 1978, 248. With permission.)

3. The anticodon and the CCA_{OH} terminus are located about 75Å apart at the opposite ends of the L and do not interact significantly with other parts of the molecule. The five bases on the 3′ side of the anticodon loop and the 3′ terminus each form a single-stranded, right-handed helical stack, but appear to be quite mobile even in the tRNA crystal.
4. The structure is dominated by hydrophobic stacking interactions which seem to constitute the main stabilizing force. Only four residues, D_{16}, D_{17}, G_{20}, and U_{47} from the extra arm, seem not to be integrated within the two helical columns of the L.
5. The junction of the L, where nearly all tertiary interactions are found, is defined by a complicated folding pattern of the D and extra arms, of the TΨC loop and of the linker nucleotides at positions 8 and 9. With one exception (the standard G_{19}-C_{56} pair), the tertiary interactions involve a series of non-Watson-Crick pairs and base-backbone interactions. Much of the tertiary hydrogen bonding between bases implicates constant residues which appear looped out in the cloverleaf; most prominent

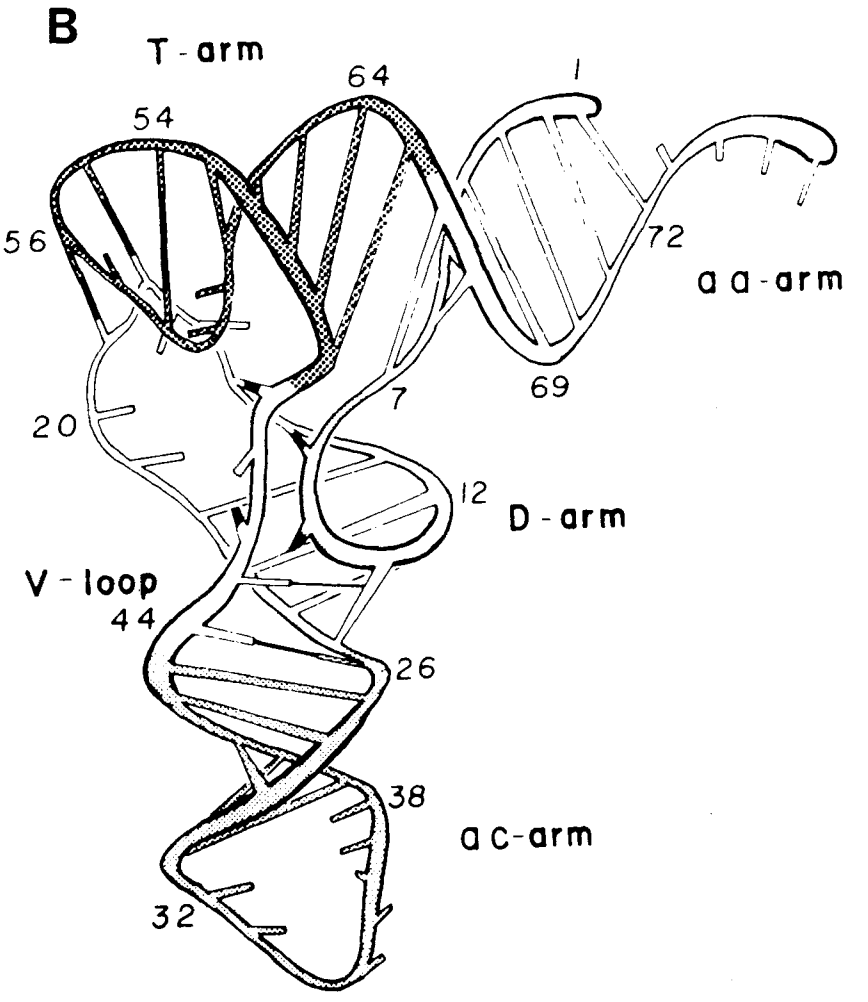


FIGURE 4B

are: the invariant pairs $G_{18} \cdot \Psi_{55}$, $G_{19} \cdot C_{56}$, $U_8 \cdot A_{14}$, $G_{15} \cdot C_{48}$ and $T_{54} \cdot m^1A_{48}$. In addition, residues G_{45}, m^7G_{46}, and A_9 contribute to form three base triples by binding each to a different base pair located in the major groove of the D stem helix.

The involvement of many of the constant residues in tertiary interactions is best understood as the result of selective pressures acting to conserve conformational parameters. This molecular design allows for each tRNA involved in polypeptide chain elongation to interact with factor Tu, ribosomal A and P sites, the CCA transferase enzyme and stringent factor, while still performing its specific functions such as aminoacylation, decoding, transcriptional regulation, and so on. It should be stressed, however, that crystallographic data give only a static picture of tRNA structure; the functioning molecule is more likely to be flexible, undergoing conformational changes in its interactions with other components of the translation machinery.

B. Scope of the Review

As alluded to previously, the sheer quantity of tRNA data available makes it impossible to analyze and evaluate all aspects of tRNA structure and function. Thus, the

following topics have been selected, derived from areas where sufficient data exist and where evolutionary hypotheses have been formulated:

1. The methodology of sequence comparison is discussed with emphasis on the validity of information tRNA can provide on biological evolution.
2. In another section, use is made of statistical methods to evaluate the evolution of tRNA in terms of mutational maps and the implications of these maps for the secondary and tertiary structure of tRNA. Only tRNA families represented by enough sequences from divergent organisms are considered in this way.
3. The significance of aminoacyl-tRNA synthetase recognition of tRNA is evaluated in terms of some published evolutionary hypotheses.
4. The distribution and evolutionary role of modified bases is analyzed.
5. An in-depth study of the crucial codon-anticodon interaction leads to hypotheses on genetic code evolution.
6. Some surprising results on mitochondrial tRNA structure and function are presented together with an analysis of T_4-coded tRNAs.

Finally, a chapter formulating or reformulating some basic questions concerning tRNA structure, function and evolution is included. Throughout the review unwarranted speculation is kept to a minimum, although some, hopefully provocative, hypotheses are proposed.

II. METHODS

A. What is a Phylogenetic Tree?

An organism's nucleic acid and protein sequences resemble those of its ancestors in some respects and differ from these ancestral sequences in other, partially random, ways. Given a set of comparable sequences from a variety of organisms, the task of phylogenetic inference is quantitatively to use sequence similarities and differences in order to hypothesize to what extent they share common ancestry and thus reconstruct their evolutionary history.[46] This history is expressed in terms of a tree diagram consisting of a set of vertices or nodes and lines or branches connecting some pairs of nodes. To define a tree, the configuration of vertices and lines must obey two topological restrictions as shown in Figure 5A: it must be simply connected and contain no closed loops. In other words, one must be able to travel from any vertex in the diagram to any other along some combination of branches, but one must not be able to return to the original point by any other route. In the biological context, the sequences being compared are related to a unique ancestral gene and, once two genes have entered distinct lines of descent (through either speciation or duplication), no further interaction between them is permitted.

In general, one point on the tree, not necessarily the vertex, is singled out as the *root,* or point of earliest time. Traveling away from this root along the branches of the tree corresponds to the passage of time. When a vertex is encountered and there is a choice of two directions in which to continue, this corresponds to an idealized speciation event. The terminal vertices correspond to present-day organisms, or rather to the data sequences pertaining to them, and the interior or nonterminal vertices represent ancestral sequences. Each branch of the tree corresponds to an ancestor-descendant relationship, and its length may be set proportional to the number of structural differences intervening between the ancestor and descendant sequences.

B. Alignment

In methods used to derive phylogenetic trees, one must have, or obtain, an appropriate alignment of the sequences as discussed in Section I and Figure 1 above. In setting up an

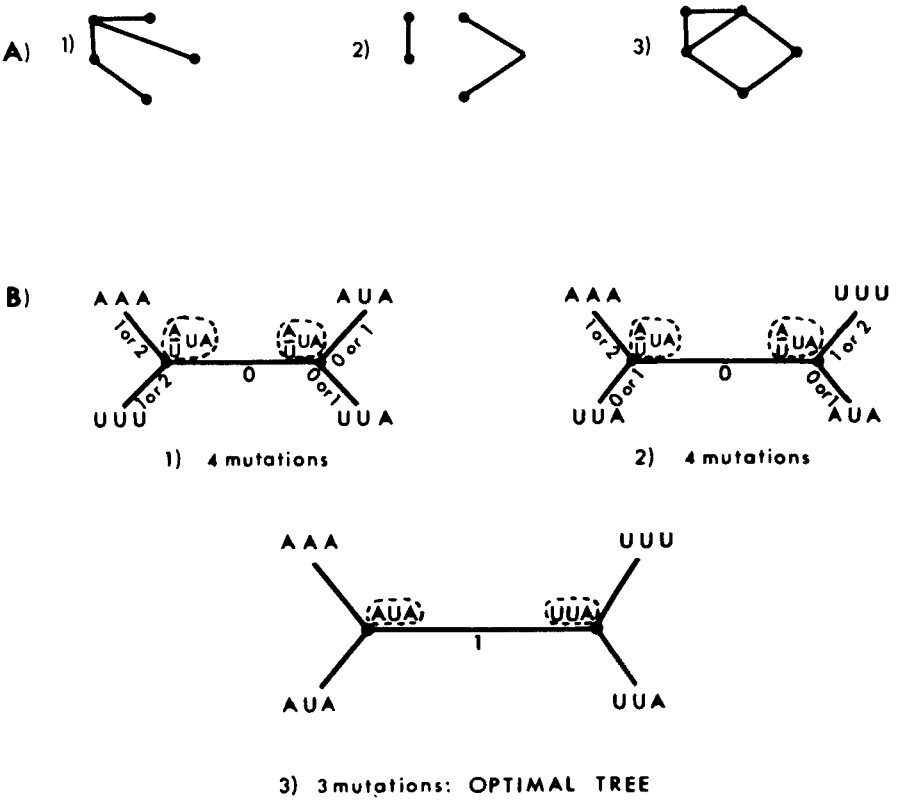


FIGURE 5. (A) Topological properties of a tree diagram. Only network (1) obeys the two essential restrictions, being simply connected and containing no loops. Network (2), not simply connected, and network (3), with closed loops, do not constitute a tree diagram. (B) Selection of the minimal tree for four sequences, AAA, UUU, AUA, and UUA. For a given tree, one selects at each node and for each position the nucleotide which leads to the minimal number of mutations when the whole network is considered. Uncertainty in the ancestral sequences, as indicated by X/Y, may lead to various estimates of the individual branch lengths, but does not modify the total mutational distance. Ancestral sequences at each node are encircled. All three possible topologies of the four sequences are given with the corresponding number of mutations implied for each. Here the topology in (3) is optimal, since it can be constructed with only three mutations.

alignment, one tries to maximize matches of identical elements from equivalent positions of different sequences. To obtain the best alignment, it is sometimes necessary to introduce gaps, or deletions, in some of the sequences which are being compared. This involves the difficult problem of postulating, in quantitative terms, the relative desirability or plausibility of inferring such a gap compared to the more prevalent type of mutation where one element is simply replaced by another. In protein studies, the alphabet of 20 different amino acids facilitates the recognition of homologous regions in different sequences, since it is exceptional that similar patterns could have arisen in two sequences purely by chance from such a large alphabet. This usually results in a unique and straightforward solution to the alignment problem. The problem is more difficult with RNA, where the alphabet contains only four different elements. In this case, it is harder to exclude the possibility that apparently homologous regions in two sequences could have been generated by chance, unless there is some additional evidence, such as secondary structure similarity, that the two regions are homologous. Thus, in 5S RNA, where until recently there has been no clear agreement on a common secondary structure,[47] a mathematically justifiable alignment could be obtained only through

difficult and sophisticated computer techniques.[16,48-50] Fortunately, tRNA sequences, by virtue of the cloverleaf base-pairing scheme and the existence of invariant residues, can be aligned according to the principle of structural equivalence and only a limited amount of ambiguity occurs in variable length regions.

C. Minimum Mutation Criterion

The goal of reconstructing the phylogenetic tree cannot usually be realized with complete confidence. Given a set of sequences, any number of trees could be postulated which satisfy the definition, and many of these may seem equally reasonable, or equally dubious, from a biological point of view. This leads to the question of how to evaluate the various evolutionary hypotheses represented by different trees from the same set of data.

Many evaluation criteria have been proposed and used, though sometimes they are only implicit in tree construction methods. Here one such criterion will be stressed, one which is frequently used and can be justified from a statistical point of view. The criterion is that of minimum mutation distance, and it may be thought of as a generalized parsimony criterion or as one which provides the most economical explanation of the given data.[16,51,52] It also generally produces "maximum likelihood" tree topologies, especially when mutation rates are not too high compared to the time span under consideration. Maximum likelihood branch lengths may also be obtained from the results of the minimum mutation approach after readjustment of the crude lengths so as to take into account parallel mutations and multiple hits which otherwise would go undetected. For this purpose, several correction procedures based on various stochastic models are currently in use.[1,53,54]

The method used by the authors of this review seeks to obtain a minimum mutation tree, that is, the tree which, among a large number of hypothetical schemes, will have the smallest total branch length, where each branch length is defined as the number of mutations necessary to transform the sequence associated with one end node of this branch to the sequence associated with the other. Such a mutational count includes base substitutions and modifications as well as deletions and insertions; in some studies, the different types of mutations are differentially weighted according to their probability of occurrence.[50] Figure 5B illustrates the basic procedure for the derivation of ancestral sequences and the choice, based on the mutational count, of the best tree.

D. An Algorithm

Though the minimum mutation criterion can be stated in a fairly straightforward way, the problem of finding or constructing the tree which satisfies this criterion is no easy task once there are more than a few sequences involved. Indeed, it is theoretically unfeasible for very large data sets belonging to the class of so-called "NP-complete" problems recognized by computer scientists as being computationally very difficult.[54a] For practical purposes, however, the situation is far from hopeless. If one is willing to accept a small risk that the tree obtained by a given method is slightly suboptimal, differing in some details from the true minimum mutation tree, relatively easy and efficient methods are available. The keys to these near-optimal methods are

1. Efficient methods for finding the truly optimal ancestral sequences for any given topology. As numerous interchanges must be attempted and rejected before finding the best tree (see below), it is essential to have a method which derives quickly these ancestral sequences for any given tree. The procedure of the authors works on the principle of dynamic programming and is quite rapid, depending only linearly on the number of ancestral nodes in the tree.[55] For each position in the sequence, it first sets up a list of possible elements at each ancestral node, N, based on the lists previously

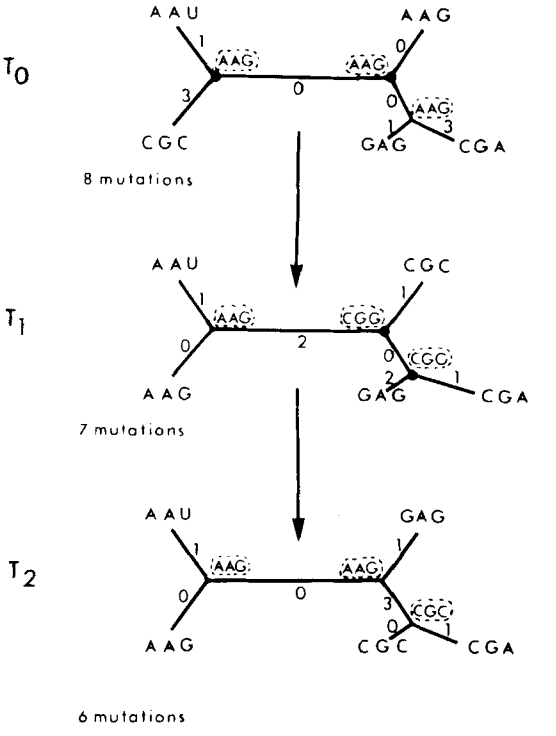
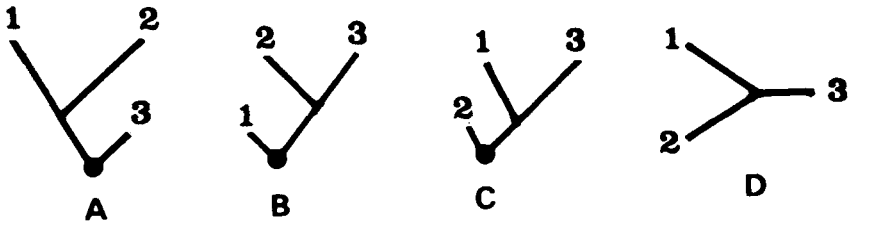


FIGURE 6. Algorithm for finding the minimal mutation tree. The example shown here is based on five data sequences: AAU, CGC, AAG, GAG and CGA. In this example, no deletion or insertion is allowed. T_0 is a tree topology chosen arbitrarily (e.g., at random), and data sequences are assigned to terminal nodes in any order. Optimal ancestral sequences (encircled) are then calculated for each interior node, as well as the mutational length of each branch. Focusing on one interior branch (indicated by heavy dots at the endpoints), the rest of the tree is decomposed into the four subtrees, namely the three isolated branches terminating in AAU, CGC and AAG, respectively, and the three-branch subtree containing GAG, CGA, and one interior node. Interchanging the AAG and CGC subtrees leads to T_1. By calculating the optimal ancestral sequences (encircled in T_1), and the mutational distance represented by each branch, it is found that T_1 is an improvement over T_0. Another subtree interchange starting for T_1 leads to further improvement, T_2, but no improvement can be made on this.

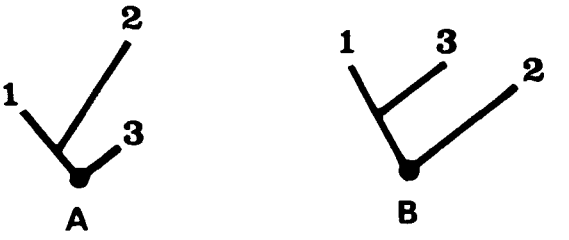
constructed for nodes intervening between N and present-day sequences. Starting at the root,* it then picks out of each nodal set, the optimal element. Repeating this procedure over all positions of the sequences permits the computation of the total mutational distance of the tree.

2. A process of local optimization to find the tree topology. The first step is to start with an arbitrary tree topology T_0, with the data sequence assigned to the terminal vertices

* Although the treeing procedure alone will not yield the point of earliest time, or root (Section II.E.), the position of the root may be arbitrarily set for computing purposes, since doing this will not change the mutational distance of the tree.



(A)



(B)

FIGURE 7. (A) Root problem. The trees A, B, and C built for the sequences 1, 2, and 3 for which a hypothetical root is indicated by a dot, possess the same topology as in D. In the absence of other criteria, the root cannot be determined. (B) Mutation rate problem. A represents here the true evolutionary history and the dot represents the root of the tree. If the branch lengths are set proportional to the number of mutations, the assumption, implicit in an unweighted hierarchical clustering procedure, of a uniform mutation rate will produce the incorrect tree B.

in any order, such as with the first tree in Figure 6. Then the optimal ancestral sequences and hence the total mutational distance are calculated for this tree as described in the previous paragraph. Now, by focusing on any interior branch of the tree, the rest of the tree may be considered as divided into four subtrees, two subtrees attached to each of the endpoints of the branch in question. The idea is to try to obtain a better tree T_1 (that is, having a smaller mutational distance) by interchanging a subtree from one endpoint with a subtree from the other endpoint. If no better tree is found by thus examining the branch, another branch is selected and examined in the same way until an improved tree T_1 is found. In the same way, one tries to improve on T_1, and so on, such that a sequence of increasingly economical trees, T_0, T_1, T_2, ... is obtained, leading to the optimal tree T_n which cannot be further improved (Figure 6). The "optimal" tree and ancestral sequences constructed by this method might appear to depend on the initial arbitrary choice of T_0. Theoretically this is true, but in practice the procedure is repeated several times for different initial choices of T_0. If all or most of the choices lead to the same local optimal T_n, one can be reasonably confident that this is the globally optimal tree, even if there is no absolute guarantee.*

* Another way of better assuring a global optimum is to allow more general subtree interchanges in the optimization procedure rather than only between subtrees at the two endpoints of the same branch.

E. The Root

There is one type of information which methods such as that described above will not normally yield, that is the point of earliest time, or root, of the tree; only the branching order of the tree network is obtained. Thus, the trees (A) to (C) in Figure 7A are all topological equivalents of the basic network (D) which results from the treeing method above. The location of the root must then be based on either the paleontological record or the assumption that the mutation rate is constant in all branches of the tree.[11,56] Although this constancy called the "molecular clock" hypothesis is probably true for eukaryotes,[11] there is very little supporting evidence in prokaryotes. Rather, the phylogenetic tree of ferredoxin would indicate that the molecular clock runs much faster for eukaryotic sequences.[57]

Most of the tRNA trees shown in this review will for purposes of convenience locate the root midway between eukaryotes and prokaryotes.

F. Application

Once a minimum mutation tree is constructed, several by-products are available. First, the branch length or mutational distance between two sequences can be obtained. If the mutational fixation rate is known and constant throughout the tree, one may calculate the approximate time lapse since the common ancestor. Conversely, knowledge of the time of divergence from fossil records, etc., permits the calculation of the mutation rate. Next, a tabulation of all positions in present-day and ancestral node sequences allows us to derive the frequency of all types of mutations, be they transitions, transversions, insertions, or deletions, and to assess the relative probability of these changes and their likelihood of fixation. Finally, the comparison of different positions in the sequence permits the discovery of "hot spots", where mutations have occurred extensively and repeatedly, and of stable regions where little change has occurred during evolution.[58] This mutational map can in turn be correlated with functional and conformational requirements of the macromolecule.

G. Matrix Methods

The minimum mutation method for tree construction is not the only way to use sequences to reconstruct phylogenies. Approaches borrowed from cluster analysis use agglomerative hierarchical grouping schemes. These are based on a matrix of mutational distances between each pair of data sequences rather than the sequences themselves. In the algorithm, the most clearly related sequences are joined to a common, recent ancestral node, the next two most closely related sequences or nodes are then joined to another, somewhat earlier node, and so on. Such methods, which are widely employed in numerical taxonomy and classification theory,[14] have the disadvantage (for the purpose of this review) in that they tend to produce a tree where the mutational distance between the tree root (earliest ancestor) and each present-day sequence is the same. This does not allow for widely divergent mutational rates in different branches of the tree so that, in a situation like that described in Figure 7A, counting only the number of differences will lead to false conclusions. Modification of these methods to alleviate this problem is, however, possible.

Other methods attempt to fit trees directly to a matrix of mutational distances between pairs of data sequences which has been corrected for parallel and superimposed mutations.[57] In this case the tree topology and the branch lengths are adjusted so as to obtain the least-squared value of differences between observed and computed distances. These methods are similar in spirit to the procedure described above and may differ according to the measure of mutational distance adopted and to the technique for actually constructing the tree. It should be noticed, however, that all matrix methods consider only the number of differences between sequences and not the sequences

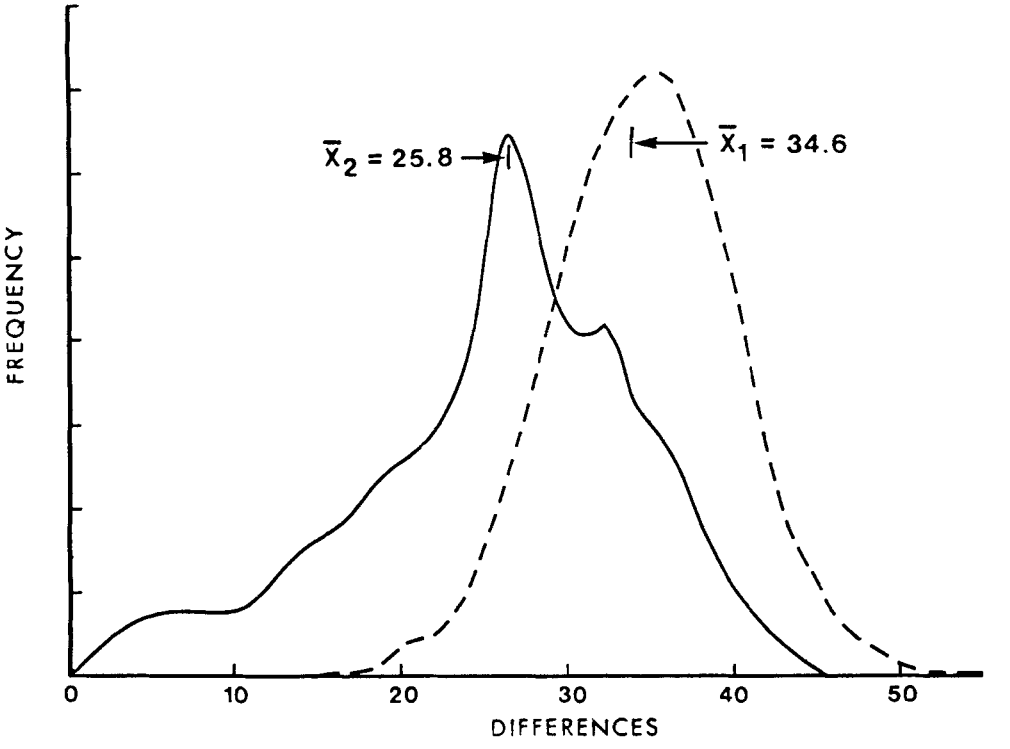


FIGURE 8. Distribution of nucleotide differences in pairwise comparisons of tRNAs. Dashed line: distribution of pairs of nonisoacceptor tRNAs. Solid line: distribution for pairs of isoacceptors.[60] For comparison purposes, the two distributions have been normalized to the same total area.

themselves. Indeed, for certain distance measures, it is impossible to construct ancestral sequences satisfying the distance relationships inferred by such methods.

H. Parallel Evolution

It could be objected that all of the above methods assume an essentially random pattern of mutation and neglect the possibility of convergent or parallel evolution.[59] Given that a common A-B branch diverged early from C, it could happen that A, after subsequent divergence from B, accumulated more mutations identical to those found in the C lineage than in B. One would then wrongly infer a common line of descent for A and C after they diverged from B. Although parallel or convergent evolution at the molecular level does sometimes occur, this is unlikely to be a general phenomenon, since it is usually found that unrelated sets of homologous proteins lead to essentially identical phylogenetic trees obtained from either sequence data or measurements of immunological cross-reactivity.[11]

I. The tRNA Problem

The algorithm for constructing trees always has an output. Thus, nonhomologous sequences or even completely arbitrary ones, such as in Figure 5B or Figure 6, can be organized into a tree network, although, of course, this may have no biological significance. In most protein studies the level of similarity is so much above random expectation that there is a tacit assumption of homology; in addition, such trees usually contain only a limited number of paralogous branches. tRNA evolutionary history, however, is undoubtedly much more complicated. Based on their common secondary

and (probably) tertiary structure, on the presence of 20 or so invariant residues, and on the sharing of several functions during translation, it is possible that all existing tRNAs diverged from a type of proto-adaptor.[1,19] If this is true, a tRNA phylogenetic tree should comprise at least 20 paralogous branches corresponding to the 20 amino acids. It must then be asked if it is still possible to reconstruct these early evolutionary events or if mutational noise, coupled with the relatively low rate of information carried by a 4-letter alphabet, allows the distinction between paralogous and orthologous events needed to infer meaningful phylogenetic relationships. Figure 8 is indicative of this noise effect: the distributions of structural differences between tRNAs specific for the same amino acid and tRNAs of different specificity overlap extensively.[60] Nonisoacceptor tRNAs in a given species in some cases share more features in common than the isoacceptors in different species, which raises the legitimate question of whether tRNAs of identical acceptor activity can be assumed *a priori* to be related. Also, changes of aminoacylation specificity could have occurred drastically affecting the homology expected among isoacceptor tRNAs. While these and other processes complicate the evolutionary record, it is nevertheless clear that the early stages of tRNA evolution, perhaps involving repeated duplications of a proto-adaptor gene, are very ancient and that a great deal of mutational activity has taken place since.

Using the limited amount of data available at the time, Holmquist et al. concluded in 1973 that phylogenetic reconstruction is very difficult if not impossible for tRNAs.[19] Nevertheless, Hasegawa has used treeing methods on an enlarged tRNA data base with some success.[61-63] He showed that a certain amount of interpretable clustering does occur, such as the sequence from initiator tRNAs or from polar amino acid specific tRNAs. Large parts of this tree, however, showed no consistent pattern.

Eigen and Winkler-Oswatitsch have also tried to reconstruct early and pre-evolutionary tRNA history by analyzing all tRNAs together or after separation by species or phyla, but not by acceptor activity.[64] Their strategy was to try to approximate directly a proto-tRNA molecule without incorporating into their algorithm information about phylogeny or homology among isoacceptor species. To do this, they chose at each sequence position, the most frequently occurring base among the present-day sequences being considered. The authors of this review consider that there are not enough data available on homologous isoacceptor families to reconstruct proto-sequences of each family though they acknowledge the desirability of this approach.

J. A Strategy

Generally speaking, it is found that most comparisons of tRNAs within the same kingdom (eukaryotic or prokaryotic) having the same anticodon give a number of differences either in the lowest five percentile of the distribution from different tRNA acceptor activities (Figure 8) or completely below this range.[60] Moreover, when various species are compared using different tRNA families, the extent of the divergence correlates well with the known phylogenetic relationship between the species. Finally, it is possible, when taking into account gene duplications that give rise to paralogous tRNA isoacceptors having the same anticodon or not, to construct a set of phylogenetic trees with essentially identical topologies. Thus, it appears that it is still possible to derive tRNA phylogenies, but such work must take account of high levels of background "noise". The problem is then one of how to pick homologous relationships out of this background. The usual treeing methods, for the reasons outlined above, are inappropriate and cannot distinguish between true homology and random similarity due to evolutionary convergence when widely divergent sequences are compared. An internal check of homology is clearly called for.

Implicit in any representation of an evolutionary tree is that the ancestral sequences

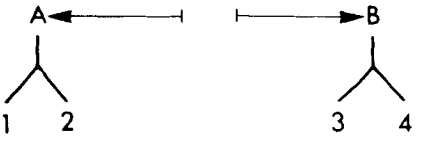


FIGURE 9. Convergence principle. A and B represent ancestral sequences independently derived from combining sequences 1 + 2 and 3 + 4, respectively. If 1, 2, 3, and 4 are homologous, then the distance between A and B should be less than the average distances of the numbered sequences.[60]

from divergent branches should become more similar as one moves toward their common root. As shown in Figure 9, if indeed a homologous relationship exists between two distinct groups of sequences, the nodal sequences A and B, *independently* constructed from 1 to 2 and 3 to 4, respectively, will exhibit more similarity than the average of the 1 to 3, 1 to 4, and 2 to 3, and 2 to 4 comparisons. In other words, the two groups should converge toward a unique ancestor, which would not be the case (or would be less the case) if they arose through independent lines of descent. Note that the term "converge" will be used here to apply to reconstructed sequences; this differs from the notion of evolutionary convergence where two or more present-day sequences are more alike than their ancestry would warrant.

The authors of this review have incorporated this general reconstructive convergence principle into a stepwise research strategy[60] somewhat different from that devised by Fitch for a related problem.[64a] First, they isolate individual groups of sequences, usually a family of isoacceptors from the same kingdom which are much more closely related than tRNAs having different acceptor activities. Such a group is considered *a priori* to be homologous, a reasonable assumption under the circumstances, and its phylogenetic tree is derived according to the minimum mutational distance criterion. Its root is located under the assumption of a more or less uniform mutation rate and this partial tree is then tentatively related through the convergence test to another group of interest, for instance the same isoacceptor family from the opposite kingdom or with a different anticodon. The test consists of comparing the similarity between the reconstructed ancestral sequences with the average similarity between the two sets of sequences used to make these reconstructions. If the ancestral sequences show significantly more similarity, the authors conclude that homology is confirmed. The new cluster thus obtained is then tentatively related to other groups of sequences through the same procedure, and so on. Of course, the feasible degree of clustering depends on the amount of available information. Although there are more than 150 known tRNA sequences, these are unevenly scattered among the 20 amino acids, the various types of anticodons, and the eukaryotic and prokaryotic kingdoms, in addition to a few organelles. The accuracy of ancestral sequence reconstruction and, subsequently, the reliability of the convergence test will depend on the number of sequences available. So far, the authors have successfully connected the prokaryotic and eukaryotic branches of $tRNA^{Phe}$ shown in Figure 10A,[58] of the initiator $tRNA^{Met}$ (Figure 10B),[58] of $tRNA^{Lys}$, of eukaryotic $tRNA^{Ser}$, and of $tRNA^{Gly}$.[60] In the last three cases, the corresponding tRNAs have more than one type of anticodon, but can all be related in a single tree. Furthermore, $tRNA^{Gly}$ and $tRNA^{Val}_{GAC}$ were shown to have shared a common ancestor (Figure 11).[60] The implications of these findings with regard to genetic code evolution will be discussed later (Section VI.).

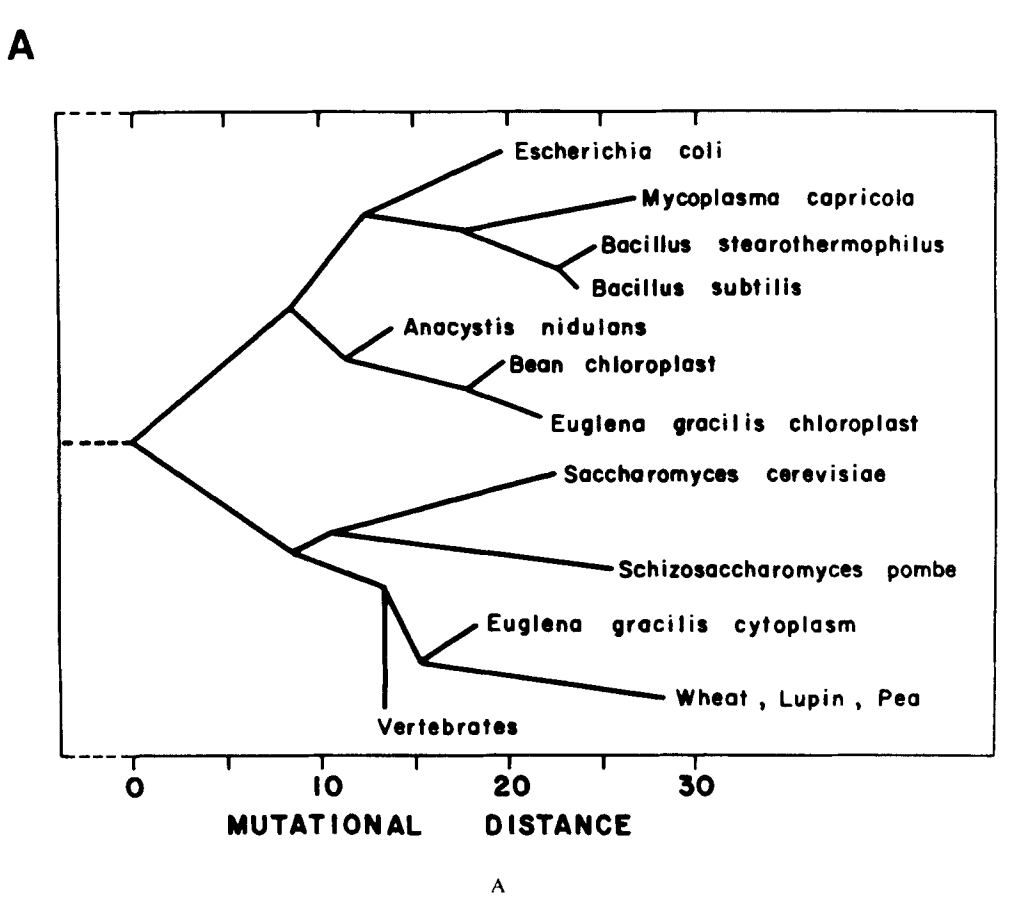


FIGURE 10. Phylogenetic trees of the (A) $tRNA^{Phe}$ and (B) $tRNA^{fMet}$ families. Trees were constructed from a minimum mutation and convergence criterion, as described in Section II. The horizontal distance is proportional to the number of mutations intervening between nodal sequences. (From LaRue, B., Cedergren, R.J., Sankoff, D., and Grosjean, H., *J. Mol. Evol.*, 1980; Cedergren, R.J., LaRue, B., Sankoff, D., Lapalme, G., and Grosjean, H., *Proc. Natl. Acad. Sci. U.S.A.*, 1980. With permission.)

At this point, it should be stressed that the proposed schemes are only partial reconstruction attempts since many homologous relationships probably have been missed for lack of sequence data. Thus, the various trees shown most likely indicate a true, but incomplete branching order. However, the glimpses now had of tRNA relationships are often sufficient to indicate which questions should be asked, in precise molecular terms, about their evolution and which sequence data, as yet unavailable, could be the most useful to resolve these questions.

III. ACTION OF THE MUTATIONAL PROCESS

A. The Significance of Sequence Data

The complexity of tRNA evolutionary history and the initial lack of sufficient sequence data have puzzled early investigators and led them to rather pessimistic conclusions.

In one of the first surveys, Dayhoff and McLaughlin built a tentative phylogenetic tree relating all of the 18 sequences known at the time, but noticed that the small number of

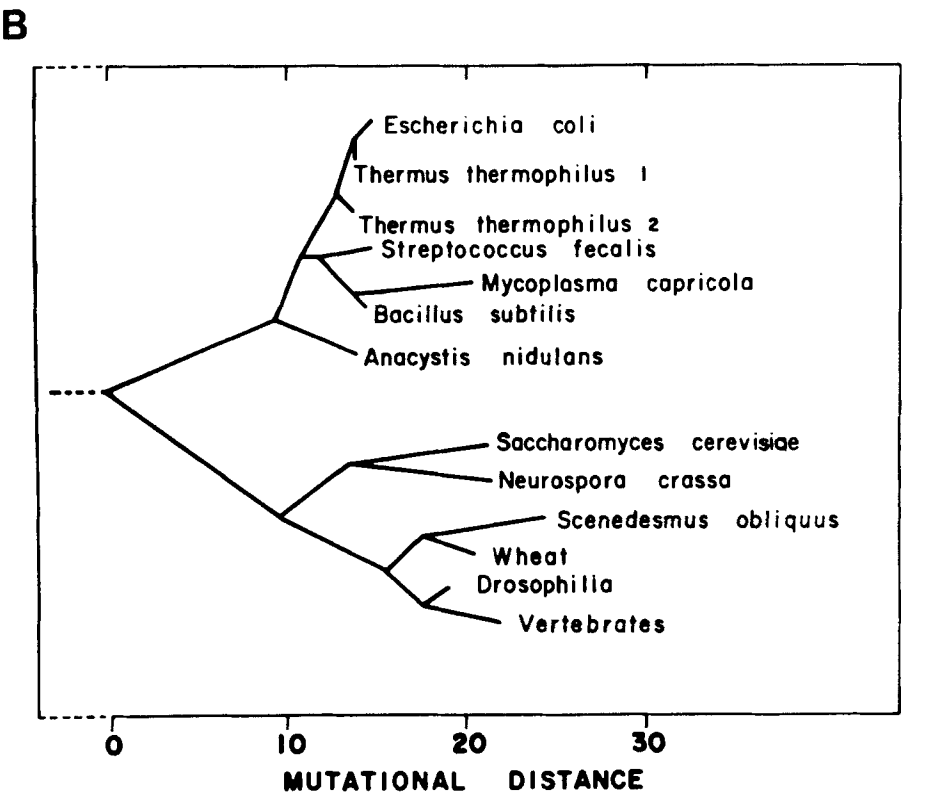


FIGURE 10B

sequence positions and the large evolutionary distances involved combined to produce a high level of uncertainty in the results,[1] The more recent investigation of Dayhoff et al. on 21 tRNA sequences from *E. coli* could only define broad families with a very imprecise order of branching.[65] In another approach, Cedergren et al. divided the tRNA molecule into a "skeleton" which consists of the >90% predictable positions and a variable "muscle".[17] Using a test based on the binomial distribution, they found a significant homology within the $tRNA^{Val}$ and $tRNA^{Ser}$ families, but none between *E. coli* and eukaryotic $tRNA^{Leu}$ which recognize codons differing at the first (5′) position.

In a systematic survey of 43 tRNA sequences, Holmquist et al. made several interesting observations.[19] They first noticed that isoacceptors from different taxa (eukaryotes and prokaryotes) diverged nearly as much as tRNAs specific for different amino acids (47.1% against 50.8% average difference), whether or not from the same organism. Next, isoacceptor tRNAs from the same species could show up to 36% difference, thus spanning a range as large as isoacceptors from the same taxon. Finally, nonhelical regions showed only 63% as much divergence as stem areas when tRNAs specific for different amino acids were compared. This was summarized by the conclusion that ". . . most of the tRNAs have diverged so widely from each other that most phylogenetically relevant information has been lost,"[19] such that no coherent phylogeny could be built. However, the theoretical maximum of 75% difference expected from total tRNA sequence randomization has not yet been reached, which suggests that current levels of divergence represent a compromise between the effects of mutation and selective pressures. In other words, tRNAs could approach a divergence limit beyond which function would be lost.

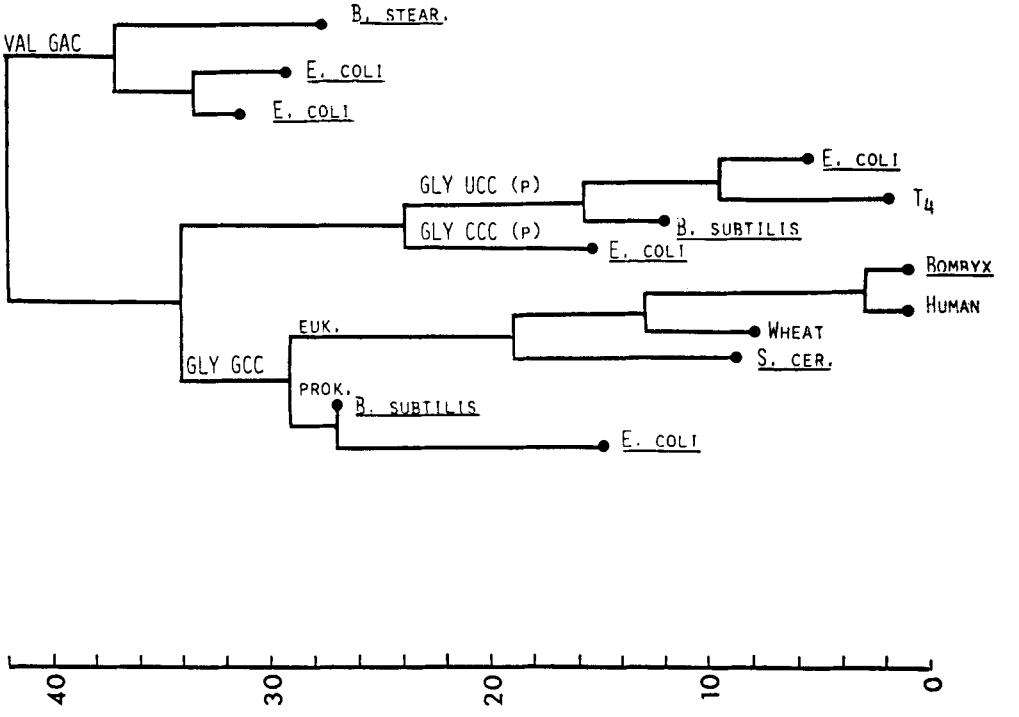


FIGURE 11. Phylogenetic tree of the Val-Gly tRNA superfamily. The relationships between tRNAs having the same anticodon were determined from a minimum mutation criterion. The branching order of these subtrees was determined from the convergence criterion. The horizontal distance is proportional to the number of structural changes between nodal sequences. The topology of the tree shows that the $tRNA^{Val}_{GAC}$ sequences diverged very early from a common glycine and valine tRNA ancestor. Next, $tRNA^{Gly}_{GCC}$ and $tRNA^{Gly}_{U/CCC}$ subfamilies diverged. (From Cedergren, R.J., LaRue, B., Sankoff, D., Lapalme, G., and Grosjean, H., *Proc. Natl. Acad. Sci. U.S.A.*, 1980. With permission.)

As new sequence data accumulate pessimism about tRNA evolutionary analysis has tempered. First, in a major work, Eigen and Winkler-Oswatitsch have shown that a single nucleotide is found more frequently than expected purely by chance at most of the positions classified as variable.[64] Presumably, this nonrandomness can be interpreted as a relic of a primitive adaptor sequence; in addition, the "ancestors" derived by this simple procedure from various isoacceptor families are more similar than current sequences, which would again suggest that mutational noise has not obliterated all evolutionary relationships. Next, as can be seen in Figure 8 (Section II.), tRNAs specific for different amino acids differ on the average at 34.6 positions while isoacceptors are usually more closely related whether or not they are from the same taxon. Thus, although it is not always possible to tell from sequence data alone (excluding of course the anticodon) whether or not two tRNAs accept the same amino acid, taxonomic relationships can often be singled out and more distant events traced back through the convergence test described in Section II. As for the discrepancies noticed by Holmquist et al.,[19] many can now be explained in terms of understanding that isoacceptors from the same species, but with different codon specificities, may be the result of long past gene duplications. The evolutionary history of an isoacceptor family may constitute a record of both speciation and specialization of the translation apparatus. Moreover, as will be shown in this Section, codon assignments and, perhaps, acceptor activities of tRNAs may not as yet be definitively settled within the genetic code, since new tRNA lines continuously appear in

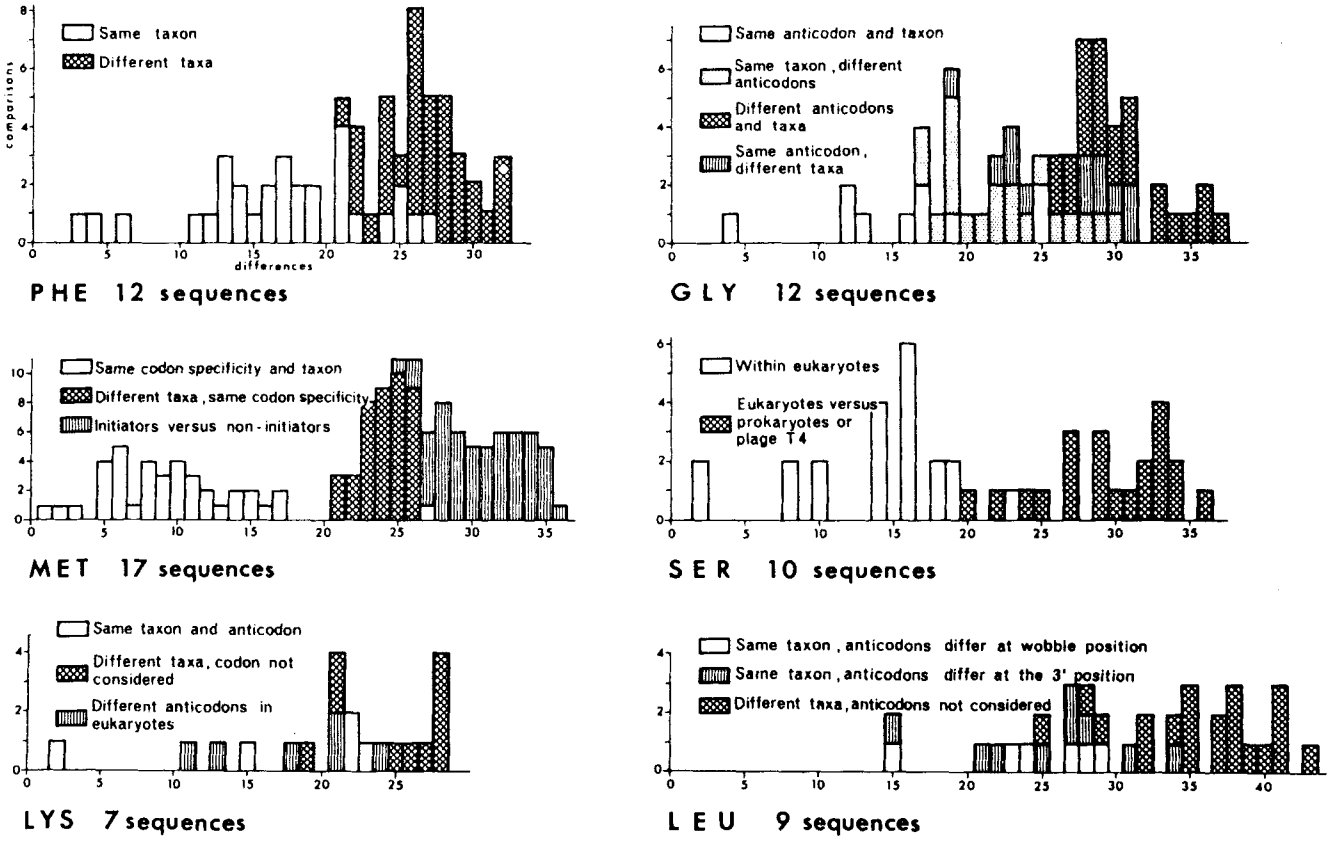


FIGURE 12. Distribution of sequence differences within isoacceptor families. The horizontal axis shows the number of structural differences between a given sequence pair, members of the indicated tRNA family. The ordinate corresponds to the number of times when this particular difference value occurs among all possible comparisons within this family. The various types of comparisons (such as between tRNAs from the same or from different taxons, and so on) have been distinguished by different shadings. The six isoacceptor tRNA families which include at least seven sequences are presented: (A) $tRNA^{Phe}$, (B) $tRNA^{Met}$, (C) $tRNA^{Lys}$, (D) $tRNA^{Gly}$, (E) $tRNA^{Ser}$, and (F) $tRNA^{Leu}$.

various phyla. These processes are quite apparent in Figure 12 which classifies several types of comparisons from the best documented isoacceptor families:

1. $tRNA^{Phe}$: a single type of anticodon, GAA, is found and most organisms investigated (assuming that eukaryotic organelles constitute distinct genetic systems) contain only one type of $tRNA^{Phe}$. Comparisons within the same taxon cover a wide range of values reflecting the more or less distant phylogenetic relationships of the organisms involved. On the other hand, comparisons of eukaryotes with prokaryotes likely refer to a single divergence event and, consequently, the distribution of the number of differences is much narrower although the numbers themselves are larger. The partial overlap of the two distributions arises mostly from the *Schizosaccharomyces pombe* sequence which followed a very peculiar evolutionary pathway (see Reference 58).
2. $tRNA^{Met}$: this family is slightly more complicated than the previous one, since it contains both initiator and noninitiator tRNAs. Comparisons fall within three distinct classes with very little overlap. The highest similarities and also the widest range of differences characterize tRNAs from the same taxon and having the same function. Next, intertaxa comparisons within either initiators or noninitiators show much more divergence but are narrowly clustered. Finally, initiators and noninitiators are the most different, whether or not they come from the same taxon: this strongly suggests that $tRNA^{fMet}$ belongs to a line whose initiator function appeared very early in evolution.
3. $tRNA^{Lys}$: two types of anticodons are known, UUU and CUU, the latter of which has been sampled so far only in eukaryotes. Except for mammalian[66] and *Drosophila*[67] $tRNA^{Lys}_{UUU}$, which differ at only two positions, eukaryotic lysine tRNAs cannot be classified neatly according to anticodon type. The evolutionary history of this family involves both speciation and gene duplication: minimum mutation trees show clearly that $tRNA^{Lys}_{UUU}$ and $tRNA^{Lys}_{CUU}$ separated only after the fungi and animal lines diverged from each other (Figure 13 and Section VI.A.).
4. $tRNA^{Gly}$: glycine is specified by the four codons of the GGX type and both eukaryotes and prokaryotes contain three kinds of $tRNA^{Gly}$, those with GCC, UCC, and CCC as anticodons. Figure 12D shows much overlap in the distributions of the four types of comparisons which can be made indicating that many taxonomic relationships have been blurred. In a recent article,[60] these data have been interpreted to mean an early separation of GCC from C/UCC anticodon-containing glycine isoacceptors, followed by the eukaryote-prokaryote divergence, and finally by the differentiation of $tRNA^{Gly}_{CCC}$ from $tRNA^{Gly}_{UCC}$ (Figure 11 and Section VI. A.).
5. $tRNA^{Ser}$: serine has six codons, four of the UCX and two of the AGC/U type. There are three prokaryotic and seven eukaryotic $tRNA^{Ser}$ sequences known which correspond to four different anticodons: UGA, CGA, IGA, and CGU. However, eukaryotic $tRNA^{Ser}$, in spite of the occurrence of all four anticodons, diverge about as much as isocoders* (such as $tRNA^{Phe}$ and each of the two $tRNA^{Met}$) from the same taxon (Figure 12). Thus, it appears likely that the various eukaryotic $tRNA^{Ser}$ diverged from each other only after the establishment of the eukaryotic line. (Figure 13).
6. $tRNA^{Leu}$: there are nine known sequences extensively scattered among the various anticodons (six in all) and taxa (data from fungi, animals, bacteria, and phage T_4). Although this extreme dispersal of data does not as yet permit the emergence of any coherent picture, it is clear that this family is very divergent even within the same taxon.

* Isoacceptor tRNAs having the same anticodon sequence.

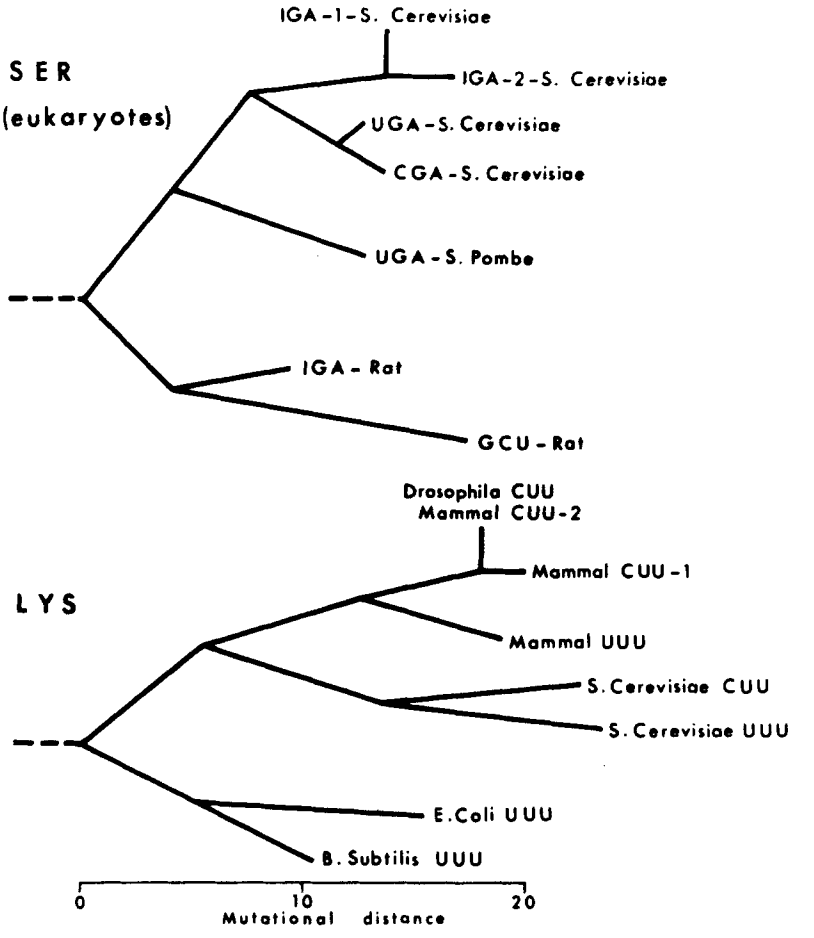


FIGURE 13. Phylogenetic trees of $tRNA^{Lys}$ and eukaryotic $tRNA^{Ser}$. These trees were derived according to the minimum mutation criterion. The number of mutations separating two adjacent nodes is given by the abcissa difference value. Each tRNA is identified by its anticodon and by the species to which it belongs. The tentative assignment of the root is based on the generally accepted taxonomy of these species.

In summary, there are a few families, such as $tRNA^{Phe}$, $tRNA^{Met}$, and the $tRNA^{Gly}$-$tRNA^{Val}_{GAC}$ cluster, where sufficient information is available to demonstrate homology and to construct phylogenetic trees. In addition, data presented here indicate that $tRNA^{Lys}$ and eukaryotic $tRNA^{Ser}$ each constitute a single homologous family.

B. Rates and Types of Mutation

Holmquist et al. compared prokaryotic and eukaryotic tRNAs specific for different amino acids and, assuming that the two taxa diverged from each other about 2×10^9 years ago, calculated a fixation rate of $2.2 \pm 0.5 \times 10^{-10}$ per nucleotide site per year.[19] If, as they argue, a saturation effect limits the extent of divergence at large evolutionary distances, the real mutational value should undoubtedly be higher. Moreover, it should be pointed out that tRNAs with different acceptor activities were compared. Since it is quite possible that the eukaryote-prokaryote divergence was preceded by a period of rapid change, where tRNAs with different specificities separated from each other, the significance of this figure is much in doubt. Recently, a fixation rate of 1.2 to 1.6×10^{-10} was directly computed from the branch lengths of the $tRNA^{Phe}$ and $tRNA^{fMet}$

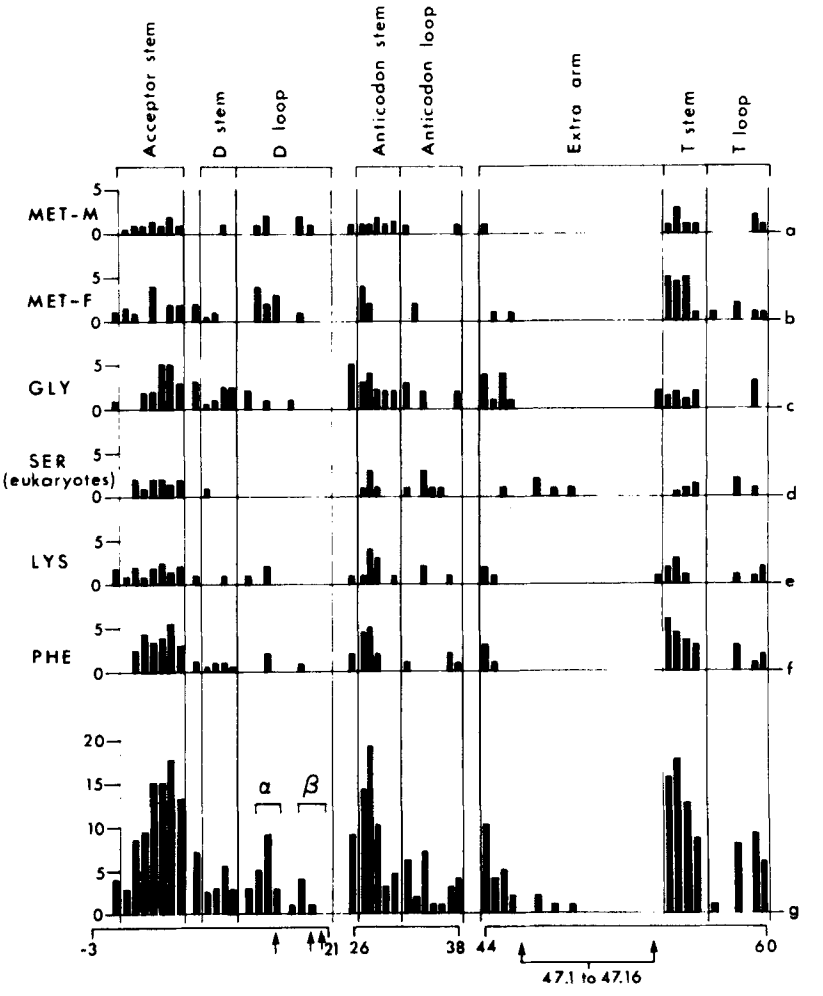


FIGURE 14. Mutation maps of six tRNA families. The phylogenetic trees of $tRNA^{Gly}$, $tRNA^{Phe}$, $tRNA^{fMet}$, $tRNA^{Lys}$, eukaryotic $tRNA^{Ser}$ (see previous section), and noninitiator $tRNA^{Met}$ (tree not shown) are each used to derive the number of recorded mutations as a function of the position in the sequence. The final panel is an accumulated map of the six tRNA families. Position numbering is as shown in Figure 2. In determining the maps, each base pair of the cloverleaf (excluding the stem of the extra arm from $tRNA^{Ser}$ in which each position was considered separately) is treated as a unit; the number of inferred mutations of a given base pair is taken as the average of the two paired positions. The mutational value is placed at the position occupied by the base closest to the 5′ terminal. The number of mutations is given in the vertical scale and the position number in the horizontal scale. Arrows indicate variable length regions of the molecule.

phylogenetic trees[58] with the assumption that the two taxa separated 1.8×10^9 years ago.[68] Since it is uncorrected for parallel mutations and multiple hits, this estimate represents a lower limit rather than a ceiling. While this figure is an average over all positions, a closer inspection of the mutation maps (see Figure 14) reveals that the various regions of the tRNA molecule differ sharply in their fixation rates. In particular, a cluster of 26 positions from the acceptor, T, and anticodon stems constitutes the best mutational targets in the molecule (Figure 15); a few other hot spots are found scattered over the rest of the sequence. Considering only these 26 positions and their average base composition, one should expect about 19 differences between random sequences. In fact, eukaryotic and prokaryotic clusters from $tRNA^{fMet}$ differ in an average of 12.6 positions

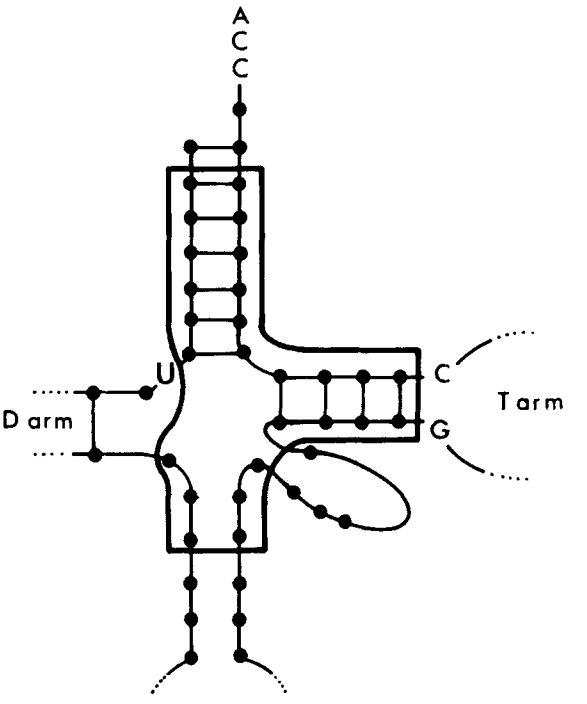


FIGURE 15. The 26 mutational hot spots in tRNA. This drawing is based on the information of Figure 14 and Section III. B. of the text.

which corresponds to about 68% randomization. Assuming a uniform fixation rate for the 26 positions, there have been 0.40 to 0.45 hits per position following the eukaryote-prokaryote divergence, which is equivalent to 2.2×10^{-10} replacements per nucleotide site per year.* However, the cluster of 26 positions has undergone extensive changes in $tRNA^{Phe}$, since no more similarity is found between prokaryotes and eukaryotes than is expected between random sequences.

Others[64] have disputed the conclusions of Holmquist et al.[19] as well. On one hand, it seems evident that mutational equilibrium has not been reached with respect to constant or near constant positions. On the other hand, it also seems clear that certain stem "hot spots" have seen repetitive cycles of mutation and have reached randomization or mutational equilibrium, at least insofar as this is possible, given functional constraints.

There is some evidence to indicate that the mutational fixation rate varies in different branches of tRNA phylogenetic trees. This has been suggested for $tRNA^{fMet}$ whose sequence is highly conserved among prokaryotes[58] and for the $tRNA^{Gly}$-$tRNA^{Val}_{GAC}$ superfamily.[60] It also appears that several tRNAs from animals or higher plants are "living fossils". For example, initiator $tRNA^{Met}$ sequences from various vertebrates ranging from bony fishes to mammals are identical;[3] mammalian and *Xenopus*[70]

* Given the fraction π of different nucleotides between homologous genes, the number $\lambda_b t$ of hits since they diverged is

$$\lambda_b t = -\frac{3}{8} \ln \left(1 - \frac{4}{3}\pi\right)$$

where λ_b is the rate of substitution per site per year and t the divergence time.[69]

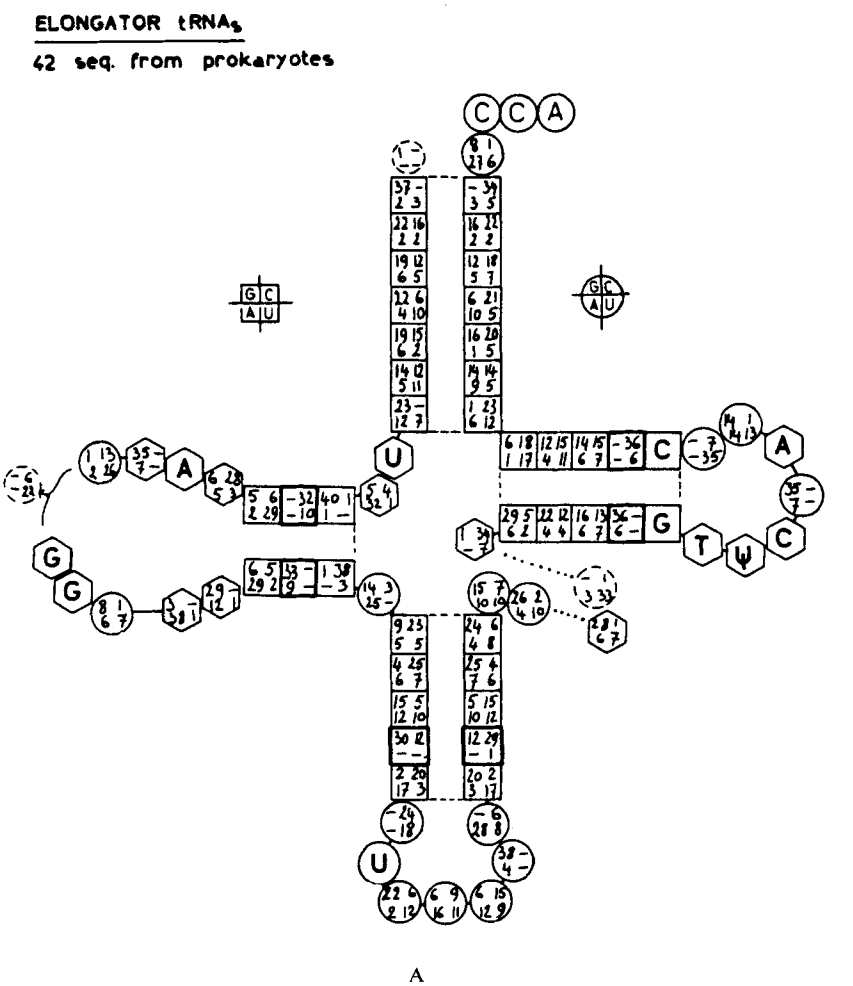


FIGURE 16. Nucleoside distribution in tRNAs. Symbols are those described in Figure 2. The number of times each of the major nucleosides is present is given at each position: G, upper left; A, lower left; C, upper right and U, lower right. (A) Cumulative distribution from 42 prokaryotic elongator tRNAs ($tRNA^{fMet}$ is therefore not considered); (B) cumulative distribution from 62 eukaryotic elongator tRNAs.

$tRNA^{Phe}$ have the same sequence which differs at only two positions from that of *Drosophila*;[71] $tRNA_2^{Lys}$ (anticodon CUU) from *Drosophila*[67] and $tRNA_{CUU}^{Lys}$ from rabbit liver[66] have the same sequence; this also applies to $tRNA^{Phe}$ and, likely, to $tRNA^{fMet}$ in mono- and dicotyledons.[72] Thus, there are instances of sequence identities between species which have diverged as long as 6×10^8 years ago, whereas approximately 12 differences are expected from the previously calculated substitution rate of 1.2 to 1.6×10^{-10} per position per year. It may be speculated that special correction mechanisms for repeated sequences, such as tRNA genes, have been developed in higher eukaryotes; this contrasts with the situation of unique sequence-coded proteins such as cytochrome *c* or the hemoglobin types, which show a much more uniform mutation rate.[11]

Very few data have been published with regard to the types of mutation found in tRNAs. Dayhoff and McLaughlin examined four closely related pairs of isoacceptors and concluded that the helical regions of the cloverleaf change 2.5 times faster than the loop areas and that A → G and C → U interchanges occur most frequently.[1] Since there

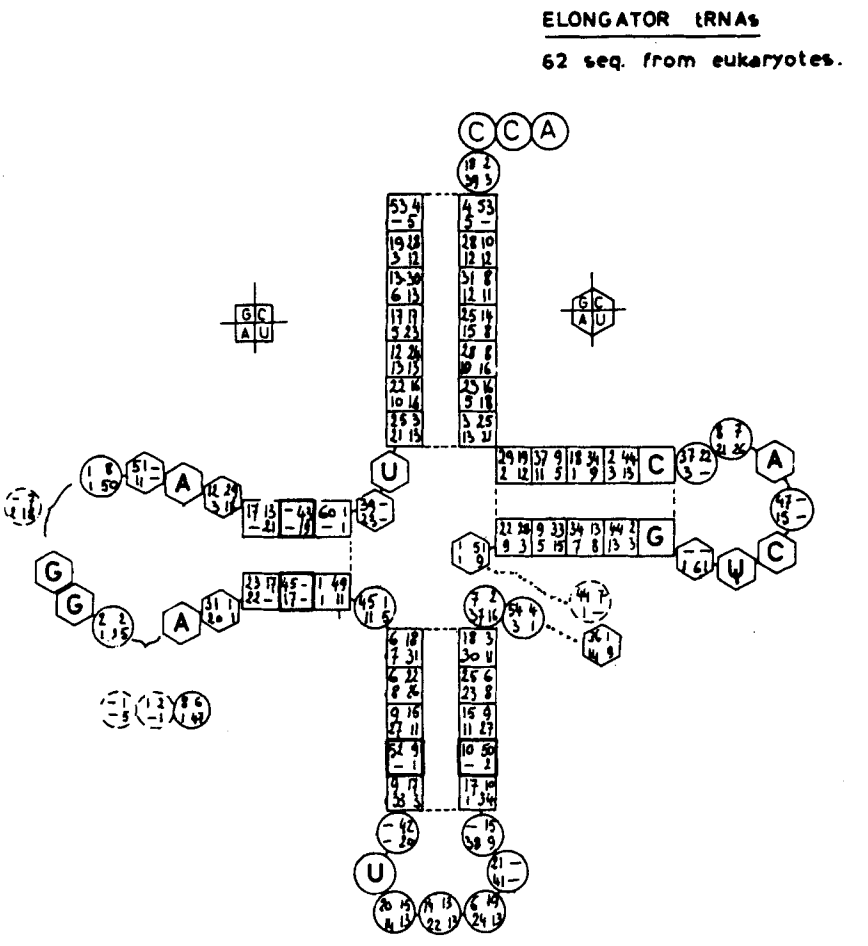


FIGURE 16B

were not at that time sufficient data to construct phylogenetic trees, the directionality of mutations could not be studied. Eigen and Winkler-Oswatitsch have compared actual sequences to the "ancestor" calculated at each position and they have concluded that C → U and A → G interchanges are quite common, while transversions are counterselected.[64] However, since this method does not involve the construction of intermediate nodal sequences, many parallel and superimposed mutations may have been missed. The availability of phylogenetic trees for $tRNA^{Phe}$ and $tRNA^{fMet}$ permits reexamination of the mutation patterns. We find that in both families G → A and C → U replacements are the most common, that each type of pyrimidine-pyrimidine and purine-purine transition tends to occur more frequently than each type of purine-pyrimidine or pyrimidine-purine transversion. Also, G and C are the most frequently replaced bases.[58] Given that over 80% of all recorded mutations occur in the stem areas, G-U and A-C are, thus, the preferred intermediates in Watson-Crick base pair interchanges, suggesting that these two pairs are not too detrimental to helix stability. Similarly, the bias in favor of replacements starting from G and C could be explained by the high G + C content of the stems. Since the high G + C content of stems and the preponderance of transitions are also found in 5S RNA,[50] it is likely that the fixation of natural mutations in all types of structural (noncoding) RNAs proceeds according to similar mechanisms (Table 2).[58]

Table 2
INFERRED MUTATIONS IN THE $tRNA^{Phe}$ AND $tRNA^{Met}$ FAMILIES

To	G	A	U	C
From				
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)	

Note: The original base is on the left, the new base is on the top. Number refers to all mutations inferred from 5S data as well as $tRNA^{Phe}$ and $tRNA^{Met}$. Numbers in parentheses are from $tRNA^{Phe}$ and $tRNA^{Met}$ only.

From LaRue, B., Cedergren, R. J., Sankoff, D., and Grosjean, H., *J. Mol. Evol.*, 14, 287, 1979. With permission.

C. Evolution and tRNA Conformation

The base-pairing scheme of the cloverleaf and the involvement of many conserved residues in tertiary bonds of yeast $tRNA^{Phe}$ most likely imply, with local variations, a common tertiary structure for all tRNAs. That the same conformation also exists in solution has been confirmed by various techniques including chemical modifications, limited nuclease digestions,[20,30] fluorescence measurements,[73] photochemical cross-linking of s^4U_8,[74-77] binding of complementary oligonucleotides[78] and NMR studies.[79]

The overall tertiary structure of tRNA is thus clearly an evolutionary invariant. It is likely that the involvement of most constant residues in tertiary interactions is the factor responsible for this structural conservatism.

Previous classifications distinguished three broad groups of tRNA according to the length of the extra arm and to the presence or absence of a Watson-Crick pair between residues 13 and 22. However, a nonstandard pair could likely be accommodated at these positions without distortion of the tertiary structure so that it appears more convenient to distinguish only tRNAs with short and long extra arms.[80] According to Brennan and Sundaralingam, tertiary interactions also fall into two classes, coincident with the previous ones, in that tRNAs with a long extra arm are devoid of base triples (although this may also be the case of a few tRNAs with a short extra arm).[81] These tertiary interactions are summarized in Table 3, to which should be added the following comments:

1. In yeast $tRNA^{Phe}$ the outside corner of the L is made of the T loop and the invariant Gs of the D loop at positions 18 and 19; that is, the pair $G_{19} \cdot C_{56}$ is made possible by a sharp turn of the polynucleotide chain after Ψ_{55} which is stacked with T_{54} and H-bonded to G_{18}; this turn is further stabilized by the $T_{54} \cdot m^1 A_{58}$ pair and by backbone interactions of Ψ_{55}. In addition, the extended conformation of the T loop is maintained by an H-bond between N^4 of G_{61} and the 5′ phosphate of C_{60}, a fact which seems to explain the constancy of the G_{53}-C_{61} base pair at the end of the T stem.
2. Linkage of the acceptor arm with the D stem requires an extension of the polynucleotide chain which is made possible by the 2′-endo pucker conformation of sugar residues 7 and 9. This allows the formation of the $U_8 \cdot A_{14}$ pair and of the $A_9 \cdot A_{23}$-U_{12} triple. The D stem region is further widened into a so-called augmented D helix formed by two other base triples which involve residues G_{45} and m^7G_{46} from the extra arm. Many alternate base triples are found in other tRNAs, but they

Table 3
TERTIARY BASE-BASE INTERACTIONS IN tRNAs

Found in Yeast $tRNA^{Phe}$[2,30]	Postulated Substitutes or Exceptions in Other tRNAs
Type A interactions: occur in most or all sequences, independently of the extra arm size[81]	
$U_8 \cdot A_{14}$, reversed Hoogsteen pair	U_{14} in Neurospora mitochondrial $tRNA^{Tyr}$
$G_{15} \cdot C_{48}$, trans base pair	$A_{15} \cdot U_{48}$, same conformation. In a few cases nonstandard A·C, G·A, G·G and U·C pairs are found
$G_{18} \cdot \Psi_{55}$, N_1 and N_2 of G_{18} H-bonded to O_4 of Ψ_{55}	$G_{18} \cdot G_{55}$ in Neurospora mitochondrial $tRNA^{fMet}$. U_{55} in eukaryotic $tRNA^{fMet}$
$G_{19} \cdot C_{56}$, Watson-Crick base pair	
$T_{54} \cdot m'A_{58}$, reversed Hoogsteen pair	U_{54} or derivatives other than T in several sequences. Position 58 is frequently not modified. A_{54} in eukaryotic $tRNA^{fMet}$ and Bombyx $tRNA^{Ala}$.[3]
Type B interactions: found in most tRNAs with short extra arm, but not in tRNAs with long extra arm[81]	
Base triples located in the large groove of the augmented *D* helix	
$m^7 G_{46} \cdot G_{22}$-C_{13}, noncoplanar pair with $m_2^2 G_{26}$ stacked on C_{25} and A_{44} on G_{43}.	A·A-U,[b] G·G-U, A·A-A, A·G-Ψ G·C-G, m'G·A-U, m'G·G-C G·G-U, U·G-C, A·G-C, C·G-C
$A_9 \cdot A_{23}$-U_{12}	Usually a purine-purine, infrequently a pyrimidine-pyrimidine or purine-pyrimidine combination in tRNAs with short extra arm. $Purine_{26}$-$pyrimidine_{44}$ pair in tRNAs with long extra arm.
$G_{45} \cdot m^2 G_{10}$-C_{25}	
$m_2^2 G_{26} \cdot A_{44}$	

[a] (-) Base pair found in the cloverleaf structure; (·) found only as a tertiary interaction.
[b] Positions of bases given in the same order as for yeast $tRNA^{Phe}$. Many rare substitutes of the base triples have been omitted.

normally do not require important adjustments of the backbone and may practically be considered isomorphous with the standard conformation.

3. In tRNAs with a long extra arm, positions 45 and 46 are part of the extra stem and thus unavailable to form base triples. It is also frequently observed in the same class of tRNAs that both positions 9 and 23 are Gs, a feature which most likely prohibits formation of the third triple. By taking account of sterical constraints and assuming $purine_{26} \cdot pyrimidine_{44}$ and $purine_{15} \cdot pyrimidine_{48}$ pairs to occur, Brennan and Sundaralingam have built a molecular model of tRNAs with a long extra arm:[81] it is then found that this arm should point away from the CCA terminus and make an angle of about 30° with the plane of the L. Otherwise, the structure is essentially similar to the one found in tRNAs with a short extra arm, since it preserves the overall geometry of the two stacked columns and most of the remaining tertiary interactions.

In conclusion, it can be stated that the base triples likely represent features which are not particularly crucial in determining tRNA conformation. This leaves space for the extra arm to bulge out of the standard structure; similarly, regions α and β of the D loop can vary in length without significantly disturbing tertiary interactions. In all, the cloverleaf and the invariant residues provide a common framework within which individual sequence variations allow for the fulfilling of tRNAs specific functions.

In a recent study which, however, excluded mitochondrial sequences, mutation maps for the $tRNA^{Phe}$ and $tRNA^{fMet}$ families were correlated with the secondary and tertiary

structure.[58] In both cases, a composite of the acceptor and T stems, of regions α and β, and of the three base-pairs (positions 27 to 29) of the anticodon stem accounts for over 80% of the total recorded mutations although it represents only about 45% of the sequence. This agrees well with the idea that base-pair interchange can occur quite freely in helical areas and that the structure of regions α and β is relatively unimportant. On the other hand, tertiary pairs described in Table 3 are strictly conserved among all phenylalanine tRNAs, except for the one involving positions 26 and 44; this latter is stacked between the anticodon and D stems and it is likely that the identity of the two bases is not crucial. The augmented D helix is nearly invariant, but one of the three base triples found in eukaryotes does not occur in prokaryotic sequences. The *Schizosaccharomyces pombe* $tRNA^{Phe}$ is the single exception to this strong conservatism:[82] eight mutations are incorporated in its augmented D helix, but replacements are made in a coordinate fashion such that the geometry of this region is not substantially modified. This may well represent a case where the disturbance introduced by a first mutation either freed the whole region from selective pressures or, on the contrary, had to be compensated for by many readjustments. Tertiary pairs are also conserved in initiator tRNAs except that $T_{54} \cdot m^1A_{58}$ is replaced by $A_{54} \cdot m^1A_{58}$ in eukaryotes; it is likely that an A-A pair can be formed, a conclusion supported by the similar backbone tracings of the T loop from yeast $tRNA^{Phe}$ and $tRNA^{fMet}$.[44] This change, however, may have a functional significance since it is thought that the $A_{54}UCG$ sequence of eukaryotic initiators interacts on the 60S ribosomal subunit with a complementary region of 5S RNA.[83] Finally, it is noticed that, of the three postulated triples, only $m^7G_{46} \cdot G_{22}$-C_{13} is strictly conserved.

In order to see whether some generalizations could be made, the previous data were compared to the mutational maps from four other families. Figure 14 shows quite a good agreement between all six maps in that the acceptor, anticodon, and T stems are usually strong mutational targets; the D stem, on the other hand, is clearly less variable. Each family has some particular features, but this could simply be the result of the partly random nature of the evolutionary process. Some features, however, are quite striking: for example, the two base pairs that are closest to the anticodon loop are, with one exception, strictly conserved within the methionine initiator, phenylalanine, lysine, and eukaryotic serine tRNA families. Recent results have shown that initiator tRNAs have a unique anticodon loop conformation which is explained not by the loop sequence itself but, more likely, by the unique sequence of the supporting stem.[45,84] We suspect that this feature in the above tRNA families is implicated in the codon-anticodon interaction. Finally, the cumulative map of Figure 14 reveals more general trends. It can be seen that there is a tendency for mutations to concentrate in the central part of the helix which comprises the acceptor and T stems (Figure 15). Also, region α mutates about three times as often as β. The tertiary pair between positions 26 and 44, which constitutes the hinge linking the anticodon and D stems, is highly variable and usually of the purine-purine type; it could well provide a weak point permitting conformational changes in tRNA. Finally, it is quite clear from the cumulative map that constant residues are well scattered among variable bases.

Figure 16 cumulates all sequence data from eukaryotic and prokaryotic elongator tRNAs, showing the base "composition" of every position. In addition to the occurrence of the well-known invariant residues, it is seen that the frequency of each of the four bases, A, C, G, or U, is not randomly distributed. However, it is often hard to tell at this point whether the peculiarities found at a given position reflect genuine structural constraints or, instead, are the result of the phylogenetic relationships between tRNA sequences. Nevertheless, one cannot help but notice the high G + C content of the stems (72% in prokaryotes, 66% in eukaryotes). Since stems as a whole are highly mutable, this

feature can be considered as a "diffuse invariant" most likely responsible for the stability of the double helical conformation. Also, previously unnoticed restrictions are very obvious at some "variable" positions. For instance, the base pair between positions 52 and 62 nearly always has a purine on the 5′ side, while A-U pairs are heavily counterselected at positions 30 and 40. Also C, U, A, and G are unfavored at positions 7, 49, 65, and 66, respectively, which altogether link the acceptor with the T stem in the tertiary structure; more generally, a purine$_{65}$-purine$_{66}$ stack (and, arising from standard pairing rules, the simultaneous pyrimidine-pyrimidine stack of residues 7 and 49) occurs less than 5% of the time. It is possible that model building could show if such a combination involves an unusual conformation of the sugar phosphate backbone in this area of the molecule. The 5′ terminus of tRNA is most frequently G, a fact best explained by the observation that a G-C pair at the end of an helix is much more stable with G on the 5′ side.[85] Moreover, it has been shown that the occurrence of a nonstandard base pair at the end of the acceptor stem in prokaryotic initiators constitutes a sufficient criterion for these tRNAs not to form a complex with elongation factor Tu.[37] Finally, the nonrandom distribution at all positions of the D stem is undoubtedly related to the involvement of this region in the formation of the base triples.[20] In all, many subtle constraints are likely to operate in such a manner that several positions will be restricted to only a few types of mutations.

IV. AMINOACYL-tRNA SYNTHETASES

Aminoacyl-tRNA synthetases catalyze the covalent attachment of amino acids to the ribose moiety of the 3′ terminal adenosine of tRNA.[86] This is a two-step reaction, involving first the synthesis of an aminoacyladenylate intermediate and next the transfer of the aminoacyl fragment to tRNA:

$$\text{ATP} + \text{amino acid} \rightleftharpoons \text{Aminoacyl-AMP} + \text{PP} \quad (1)$$

$$\text{Aminoacyl-AMP} + \text{tRNA} \rightleftharpoons \text{Aminoacyl-tRNA} + \text{AMP} \quad (2)$$

The enzyme must exercise selectivity both at the tRNA and amino acid levels, otherwise the improper matching of tRNA and amino acid would produce errors in the protein sequence. It is currently estimated that such mistakes occur about once in ten thousand.[87] The study of synthetases has provided information on two hydrolytic corrective mechanisms: the first involving the aminoacyladenylates, the second, noncognate aminoacylated tRNAs.

In prokaryotes, there is usually a single synthetase of each amino acid; eukaryotes, on the other hand, may have up to three distinct sets of synthetases, one each from the cytoplasm, chloroplast, and mitochondrion.[88] Due to their relationship during translation, the tRNAs and the synthetases may well have followed parallel evolutionary pathways. Thus, the authors propose to evaluate the available information about aminoacyl-tRNA synthetases with regard to evolutionary hypotheses here.

A. Structure of Aminoacyl-tRNA Synthetases

The synthetase group of enzymes exhibits a considerable amount of structural heterogeneity.[86] In *E. coli,* for instance, molecular weights range from 59,000 to 227,000; the subunit structure is characterized by either α, α_2, α_4, or $\alpha_2\beta_2$ structures. Only one complete sequence, that of *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase,[89] has been reported so far although other work is now in progress. Several synthetases show internal sequence repeats; these repeats have been found in mono-, di-, and tetrameric enzymes.[90-93] In some instances, it is clear that the repeat units are widely spaced, thus suggesting the polypeptide chain is the product of a gene duplication-fusion

Table 4
AMINOACYL-tRNA SYNTHETASES AND DISCRIMINATOR RELATION

Codon group	Amino acid	Enzyme quaternary structure	2′/3′ charging specificity	Cognate tRNA discriminator
XUX	Leu	α	2′	A (A)
	Ile	α	2′	A (A)
	Val	α	2′	A (A)
	Met	α_2	2′	A (A)
	Phe	$\alpha_2\beta_2$	2′	A (A)
XCX	Ser	α_2	3′	G (G)
	Pro	α_2	3′	— (C, A)
	Thr	α_2	3′	A (A)
	Ala	α	3′	A (A)
XAX	Tyr	α_2	2′, 3′	A (A)
	His	α_2	3′	C
	Gln	α	?	G
	Asn	—	2′, 3′	G
	Lys	α_2	3′	A (G, U)
	Asp	α	3′	G (G)
	Glu	α	2′	G (G)
XGX	Cys	α	2′, 3′	U (U)
	Trp	α_2	2′, 3′	G (A)
	Arg	α	2′	A (G, U)
	Ser	α_2	3′	G (G)
	Gly	$\alpha_2\beta_2$	3′	U (A)

Note: The relation of codon group, amino acid and some properties of tRNA and their cognate synthetases. The discriminator site is the 4th base from the 3′ terminus. Eukaryotic discriminator is indicated in parentheses.

Adapted from Schimmel, P. R. and Söll, D., *Annu. Rev. Biochem.*, 48, 601, 1979; Wetzel, R., *Origins Life*, 9, 39, 1978.

cycle.[90] However, such internal duplications do not necessarily correspond to two functional sites per subunit.[86] It can be hypothesized that the actual set of 20 synthetases is derived, through a series of duplications, from a small number of ancestral enzymes which coexisted with more primitive versions of the genetic code. This was indeed suggested for three *E. coli* synthetases, based on their mechanistic properties and subunit structures, but a definite proof must await sequence data.[94]

B. Comparative Aspects of Synthetase Recognition

Recent results have progressively dispelled the initial, perhaps naïve, view that the synthetase recognition of tRNA involves a linear nucleotide sequence. Rather, as detailed in recent reviews, these enzymes seem first to bind quite nonspecifically to three contact points along the diagonal formed by the inside part of the L-shaped tRNA structure; next, a more subtle and specific adjustment is made in order to bring the 3′ terminal adenosine into the catalytic center.[86,95-97]

Crothers et al. were the first to point out the importance of the fourth base from the 3′ end and to show that there is generally a correlation between the base present at this "discriminator" site and the amino acid charged (Table 4).[100] Moreover, other workers observed that noncognate tRNA acylation usually occurs among groups of tRNA sharing the same discriminator site.[98,99,101] Another pertinent result was obtained through

substitution of the 3′ terminal adenosine by a 2′ or 3′ desoxyadenosine derivative; aminoacylation experiments with these derivatives have shown that most synthetses have either a 2′ or a 3′ charging specificity.[102] Investigations with prokaryotic and eukaryotic organisms have also shown that this specificity is conserved over the whole phylogenetic spectrum.[102-104] Wetzel has noticed that four amino acids from the hydrophobic cluster — leucine, valine, isoleucine, and methionine — correspond to a single discriminator site and charging specificity.[101] Their synthetases have quite similar subunit structures, as well. He has proposed that amino acids whose codons have a central U are late arrivals in the genetic code and that their synthetases, in addition to their common mechanistic features, should be phylogenetically related. On the other hand, synthetases whose amino acids have codons with a central purine show much more diversity and could be related to an early, incomplete version of the genetic code. Although quite speculative at this time, that scheme could be specifically tested as tRNA and synthetase sequence data are made available. Supporting this proposal, valine, isoleucine, and methionine tRNAs have been found to cluster on the same branch of a phylogenetic tree derived by Schwartz, et al. from *E. coli* sequences.[65]

C. One Synthetase: Two Non-isoaccepting tRNAs

One should expect, as normally found, that there is at least one synthetase for each amino acid, allowing thus an unequivocal matching between amino acid and tRNA. However, it has been shown that *B. subtilis* is devoid of a glutaminyl-tRNA synthetase. Instead, a monomeric enzyme, 65,000 in molecular weight, acylates both $tRNA^{Glu}$ and $tRNA^{Gln}$ with glutamic acid.[105] After, Glu-$tRNA^{Gln}$ is rapidly amidated to Gln-$tRNA^{Gln}$. When tested against *E. Coli* tRNAs, this synthetase acylates $tRNA^{Gln}$ normally, but $tRNA^{Glu}$ very inefficiently. In a preliminary survey of a few other prokaryotes, Lapointe *et al.* have found that *Clostridium pasteurianum,* usually thought to be an archaic form, contains both a glutamyl- and a glutaminyl-tRNA synthetase.[105] It is also possible that one of the two reported $tRNA^{Glu}$ from yeast mitochondria is in fact a $tRNA^{Gln}$ which can be acylated by glutamate.[106]

Although there is still too little evidence for a clear choice between them, two evolutionary schemes could be invoked to explain this phenomenon. Either a primitive translation system possessed a glutamate-specific synthetase and a Glu-$tRNA^{Gln}$ amidase, or both glutamyl- and glutaminyl-tRNA synthetases were already present. In the latter hypothesis, further developments of the translation apparatus in organisms such as *B. subtilis* could have led to a dual tRNA specificity for the glutamate-charging enzyme and the emergence of an amidase; ultimately, the unnecessary glutaminyl-tRNA synthetase could be lost. If this scenario is correct, a systematic survey should unveil the existence of organisms where all three enzymes are functional.

V. MODIFIED NUCLEOSIDES OF tRNA

One of the most distinctive structural features of tRNA is the presence of a significant proportion of modified nucleosides derived from the four parents, A, C, U, and G. All modifications are posttranscriptional — they take place after synthesis of the macromolecules during the processing of maturation of tRNA.[107,108] Only the guanine derivative Q found in the first anticodon position (and probably Ψ as well) is inserted into the polynucleotide backbone after the original base, guanine, found in the primary transcript, has been enzymatically removed.[109,110] No other modifications are known to involve a depurination or a depyrimidination step, although i^6A can apparently be enzymatically removed from tRNA.[111]

Some 50 modified nucleosides have been isolated and structurally identified. Some modifications are considered to be complex, while others are simple. The "simple"

Table 5
MODIFIED NUCLEOSIDES

Abbreviation	Nucleoside	Position	Distribution
	Adenosine Derivatives		
I	Inosine	34	E, P
m^1I	1-Methylinosine	37	E
m^1A	1-Methyladenosine	22, 58	E, P
m^2A	2-Methyladenosine	37	P
m^6A	N^6-methyladenosine	37	P
i^6A	N^6-isopentenyladenosine	37	E, P
t^6A	N^6-threonylcarbamoyladenosine	37	E, P, M
ms^2i^6A	2-Methylthio-N^6-isopentenyladenosine	37	P
ms^2t^6A	2-Methylthio-N^6-threonylcarbomoyladenosine	37	E
	Cytidine Derivatives		
m^3C	N^3-methylcytidine	32	E
m^5C	5-Methylcytidine	[a]	E
Cm	2′-*O*-methylcytidine	32	E, P
ac^4C	N^4-acetylcytidine	12, 34	E, P
s^2C	2-Thiocytidine	32	P
s^2m^5C	2-Thio-5-methylcytidine		
	Guanosine Derivatives		
m^1G	1-Methylguanosine	9, 37	E, P, M
m^2G	2-Methylguanosine	6, 9, 26	E, M
m_2^2G	2-Dimethylguanosine	26	E, M
m^7G	7-Methylguanosine	46	E, P
Gm	2′-*O*-methylguanosine	18, 19, 34	E, P
Y^W	Wybutosine and derivatives	37	E
Q	Queosine and derivatives	34	E, P
	Uridine Derivatives		
Ψ	Pseudouridine	[a]	E, P, M
T	Thymidine	54	E, P, M
Um	2′-*O*-methyluridine	4, 32	E, P
Ψ_m	2′-*O*-methylpseudouridine	39	E
D	Dihydrouridine	[a]	E, P, M
s^4U	4-Thiouridine	8, 9	P
s^2m^5U	2-Thio-5-methyluridine	54	P
Tm	2′-*O*-methylthymidine	54	E
V	Uridine-5-oxyacetic acid	34	P
s^2cm^5U	2-Thio-5-carboxymethyluridine		P
mam^5s^2U	2-Thio-5-methylaminomethyl-uridine	34	P
$4abu^3U$	3-(3-Amino-3-carboxyproply)uridine	47	P
s^2mcm^5U	2-Thio-5-(methoxycarbonylmethyl)-uridine	34	P, M
mo^5U	5-Methoxyuridine	34	P
mcm^5U	5-Methoxycarbonylmethyluridine	34	P

Note: E = eukaryotes; P = prokaryotes; M = mitochondria.

[a] Widely distributed.

Adapted from Sprinzl, M., Grueter, F., Spelzhaus, A., and Gauss, D. H., *Nucl. Ac. Res.*, 8, r1, 1980.

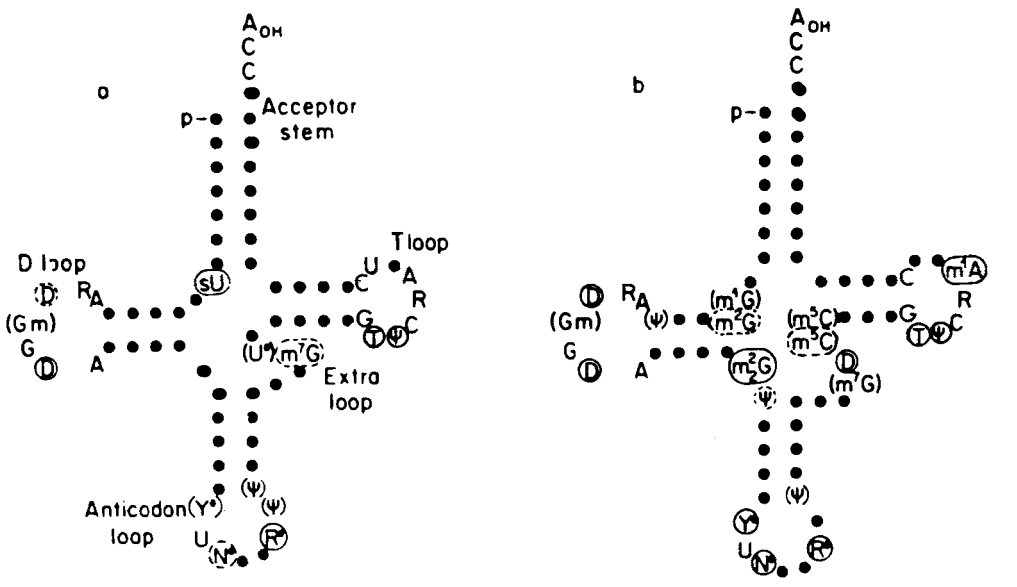


FIGURE 17. Common locations of modified nucleosides. Modified nucleosides occurring in most tRNAs are encircled; those found in about 50% are in dashed lines and those found in one third or less of known sequences are in brackets. N = unspecified nucleoside, R = purine nucleoside, Y = pyrimidine nucleoside, and U = various uridine derivatives. An asterisk indicates a modification. (a) prokaryotic tRNAs, mostly from *Escherichia coli;* (b) eukaryotic tRNAs, mostly from yeast. (Adapted from Feldman, M.Y., *Progr. Biophys. Mol. Biol.*, 23, 227, 1979. With permission.)

category is populated by nucleosides produced by reactions such as: base or ribose methylations, reduction, thiolation, etc. The majority of the complex, hypermodified nucleosides, whose biosynthesis frequently requires the addition of side chains, are found either at the first position of the anticodon (wobble position) or adjacent to the 3′ side of the anticodon.

A. Distribution

Table 5 gives a list of modified nucleosides and indicates that most modifications are either completely specific for a unique position or a limited number of positions in tRNA. It is likely that many modifying enzymes are position and base specific (Figure 17). In contrast to this general rule, D and Ψ are found in a variety of positions; the former, however, is found mostly in the D loop and occasionally at position 47 of the extra arm. In this broader sense, it can be considered to be a region-specific modification. Pseudouridine, on the other hand, is part of the invarient TΨC sequence, but it occurs also at a number of other sites in tRNA molecules, particularly in the anticodon arm. In *Salmonella typhimurium* $tRNA^{His}$ (and quite likely in other tRNAs), it has been shown that different enzymes are involved in the modification of U to Ψ in the various positions of the sequence,[113] consistent with the notion of position specificity of modification enzymes.

Some theoretical work has attempted to deduce the structural elements of tRNA involved in the recognition by the modifying enzymes. In two studies using 43 and 60 tRNA sequences, respectively,[114,115] some structural correlations were made; none of these have been confirmed in the more than 60 new sequences determined since these publications.[3] One of the problems in this type of study is to distinguish between what is a necessary and what is a sufficient condition for modification. An example can be drawn from the t^6A modification in *E. coli* tRNAs. Here it is clear that $U_{36} + A_{37}$ are necessary

Table 6
CORRELATIONS BETWEEN MODIFIED BASES AND CELLULAR METABOLISM

tRNA	Source	Modification	Interpretation	Ref.
$tRNA^{Phe}_{GAA}$	*Escherichia coli*	D, Ψ, s^2i^6A, 4 abu^3U	Modifications affected by amino acid starvation in relaxed control mutants	122
$tRNA^{Lys}$	Mouse tumor	Tm, Ψ	Present in rapidly dividing cells	123
$tRNA^{Phe}$	Mouse tumor	Y^W	Lacking in rapidly dividing cells	124
$tRNA^{Phe}$	*Rhodopseudomonas spheroides*	m_2^2G	Synthesis in photosynthetic mode	125
$tRNA^{Asn}$	Rat hepatoma	Q	Absence in tumor tissues	126
$tRNA^{Tyr}$	*Bacillus subtilis*	i^6A	Presence in place of ms^2i^6A in sporulating cells	127
$tRNA^{Tyr}$; $tRNA^{Asp}$	Chick oviduct	Q	Different Q levels in estrogen-treated and control cells	128
$tRNA^{Tyr}$; $tRNA^{Asn}$	*Drosophila*	Q	Related to development	129
$tRNA^{Asn}$	Dictyostelium	Q	Starvation leading to sporulation concomitant with Q appearance	130
tRNA	Animal tumors	Methylation	Increased methylation in tumors	131

for modification; it is not sufficient however, since $tRNA^{fMet}$ has a U_{36}-A_{37} sequence and the A_{37} is not modified.[3] A suppressor $tRNA^{Gly}_{UCU}$ does have the t^6A modification, even though the wild-type $tRNA^{Gly}_{UCC}$ does not.[116] Further work along this line will have to take into account the demonstrated presence or absence of homologous modifying enzymes in a given organism.

In spite of the preceding examples, there is a rather clear correlation to be made concerning the modification of nucleosides at the position 3′ adjacent to the anticodon. It is believed that these modifications are important for the stablization of the codon-anticodon complex (see Section VI. B.)

B. Significance of Modified Nucleosides

Whereas the existence of modified nucleosides has been known for some time, their role in tRNA function and evolution is obscure; a major exception is the case of Ψ_{38} and Ψ_{39} whose absence in $tRNA^{His}$ from *S. typhimurium* results directly in the constitutive expression of the HIS operon.[113,117] The observation that $tRNA^{Trp}$ suppression of the UGA nonsense codon results from a base change far from the anticodon loop[118] does, however, permit speculation that modification of nucleosides in other parts of the molecule could affect codon-anticodon interactions (translational control?)[119–121] Summarized in Table 6 are a number of correlations between cellular metabolism and changes in modified nucleoside complements; their evolutionary significance, however, is not clear at the present time.

Since there is little function to which modified bases can be related except for those considered in Section VI, an evolutionary analysis of modifications is premature. There is, however, a problem of a practical nature when tRNA sequences are quantitatively compared: how to deal with modifications? It is indeed possible that real phylogenetic information can be gleaned from these modifications, although in most methods they are

not considered. To support this idea, we note that V is a modification found only in *E. coli*;[3] D-loop m^1 A in *Bacillus*;[132] Y in some eukaryotes,[3] etc. Moreover, it has long been recognized that cellular complexity was directly related to the number of modified nucleosides in tRNA. Recently, Ninio suggested that tRNA modifications could be used as a taxonomic index.[114,133] In order to evaluate this hypothesis, we have reexamined the eukaryotic $tRNA^{Phe}$ tree. As noted before, usual treeing methods which do not use a convergence criterion show *Euglena* to branch close to mammals.[60] The authors thought that by increasing the importance of nucleoside modification in calculating taxonomic difference, *Euglena* might then branch closer to either plants or *Protista*. In order to do so, they arbitrarily assigned a weight of three to a change of nucleoside modification while keeping a weight of one for every nucleoside replacement during calculation of the difference matrix of eukaryotic $tRNA^{Phe}$. This procedure does not change the topology of the $tRNA^{Phe}$ tree and, particularly, the branching of the *Euglena* sequence. Although this single test is hardly conclusive as far as the hypothesis is concerned, it indicates that if the general pattern of nucleoside modifications is phylogenetically significant, it is not obvious in what manner these data can be used.

A variation of Ninio's hypothesis was recently put forward by Kubli.[134] His proposal, which is of major theoretical interest, correlates modified nucleosides to the evolution of tRNA transcribed from a multigene system. Since it is thought that repeated genes are maintained in almost identical copies by correcting mechanisms, one would not expect these tRNA genes to evolve to new sequences very rapidly. Indeed, there is good evidence (see Section III.) that tRNA genes are almost identical within both the plant and animal kingdoms. Nucleoside modifications, therefore, could serve the purpose of providing structural change in the evolving system. Mutations in structural genes for modifying enzymes would be analogous to pleiotropic mutations, since all copies of multigene products would be affected.

Although animal and plant genes evolve slowly, tRNAs from the major groups of eukaryotes, animals, plants, and fungi, are, when related to a time scale, at least as divergent as the various prokaryotic classes. Even though nucleoside modification data of tRNA families in plants and animals are sparse, the nature of modifications does not seem abnormally different. An outstanding exception is that of Y-base which was previously thought to be present in all eukaryotic $tRNA^{Phe}$, but is now known to be absent in insect $tRNA^{Phe}$.[70,71]

Even if base modification now helps to provide structural change in eukaryotic tRNAs, this cannot be the explanation for the origin of the modification process, since the ancestral organisms could have no prior "knowledge" of its eventual use. Rather the evolutionary process is opportunistic and invents new roles for existing structures. In this same way, modified nucleoside as well as tRNA roles have probably increased during evolution. One has the strong impression that the original role for modified nucleosides is yet to be found.

VI. EVOLUTIONARY ASPECTS OF THE CODON-ANTICODON INTERACTION

A. The Origin of tRNA Isoacceptor Species

The degeneracy of the genetic code is such that most of the 20 amino acids are specified by more than one code word. The apparent difficulty in having a given amino acid specified by different codons was resolved when it was realized not only that a single tRNA species may sometimes recognize several codons (wobble, Section VI.C.), but also that translation components can include several distinct tRNA species (isoacceptors) for the same amino acid.[135] Collectively, these two circumstances allow reading of all codons for a given amino acid (up to six).

Various isoacceptor tRNA species owe their existence to one of two phenomena. First, isoacceptor species can originate from different modified nucleoside levels in tRNAs.[71] Alternatively, isoacceptor tRNAs may arise from differences in the primary, unmodified structure,[3] these isoacceptors would presumably be the product of heterogenous structural genes for a given tRNA family. Here we will examine the latter class of isoacceptors, that concerned with differing primary sequences, in order to determine what is the function and the origin of this structural heterogeneity.

In some cases, the number of structural differences between isoacceptor species is so small that only a recent gene duplication followed by divergence could be a reasonable explanation. The two initiator tRNAs from *E. coli* or *Thermus thermophilus*, the two mammalian $tRNA^{Lys}_{CUU}$, the two $tRNA^{Ser}_{IGA}$ and $tRNA^{Ser}_{C/UGA}$ from yeast are but a few examples of isoacceptors differing at four positions or less. In most of these cases, the isoacceptor species share the same anticodon and base modifications. At the other extreme, $tRNA^{Arg}_{UCU}$ and $tRNA^{Arg}_{ICG}$ from yeast or $tRNA^{Ser}_{UGA}$ and $tRNA^{Ser}_{GCU}$ from *E. coli* differ at more than 20 positions; a rigorous proof of homology in these two cases will be needed before they can be related to a common ancestor through gene duplication. Finally, the eukaryotic $tRNA^{Lys}_{UUU}$ and $tRNA^{Lys}_{CUU}$, the $tRNA^{Gly}_{CCC}$ and $tRNA^{Gly}_{UCC}$ from *E. coli*, or the $tRNA^{Ser}_{IGA}$ and $tRNA^{Ser}_{U/CGA}$ from yeast represent intermediate levels of divergence. It is not the authors' purpose to review all the instances of probable gene duplications which can be inferred from the sequence data, but rather to focus on three well-documented families, namely $tRNA^{Lys}$, eukaryotic $tRNA^{Ser}$ and the $tRNA^{Val}_{GAC}$-$tRNA^{Gly}$ cluster.

1. tRNA^{Lys}

There are eight known $tRNA^{Lys}$ sequences, two from prokaryotes in which only UUU is found as anticodon, although another anticodon type does exist, and six from eukaryotes which represent the two types of anticodons, UUU and CUU. Convergence of eukaryotic and prokaryotic sequences has been found (the authors, unpublished data), which allows the construction of a single phylogenetic tree. The best topology, as shown in Figure 3 involve first the separation of fungi and animals followed by a differentiation of both lines according to anticodon type. The alternate arrangement where the two types of anticodons appear at the beginning of eukaryotic history is quite unacceptable, since it requires nine additional mutations and lowers the convergence of ancestral sequences. Thus, we have a clear example of parallel evolution at the molecular level in that two divergent lines independently developed tRNAs with similar functional features.

2. Eukaryotic tRNA^{Ser}

Of the 11 known sequences of serine tRNAs, 4 were obtained from bacteria, phage, or mitochondria, but they are not clearly related to eukaryotic $tRNA^{Ser}$. Within this latter group, however, there is much less divergence than is usually found between prokaryotes and eukaryotes; consequently, the eukaryotic $tRNA^{Ser}$ is treated as a single homologous family. As for $tRNA^{Ser}$, a unique topology, shown in Figure 13, is clearly optimal. It should be noted that the four anticodon types represented (IGA, UGA, CGA, and GCU) collectively allow the recognition of all six serine codons. From the four yeast sequences of the XGA anticodon group, an evolutionary model of increasing codon specificity can be deduced, where the wobble position is first differentiated into purine and pyrimidine branches, the latter being subsequently divided into the C- and U-containing anticodons. In spite of the fact that the two serine tRNAs from rat differ at all three positions of the anticodon, they are phylogenetically related by the topology shown. One must stress that the root of this tree cannot be unambiguously assigned; in our scheme, the root divides the tree into two approximately equal parts, but one can easily conceive of its location on the $tRNA^{Ser}_{GCU}$ branch, if very uneven mutation rates exist in different branches. More

sequence data, both from yeast $tRNA^{Ser}_{GCU}$ and from rat $tRNA^{Ser}_{U/CGA}$ are needed to resolve this point. Nevertheless, it is clear that anticodon changes have taken place quite freely during the evolution of eukaryotic $tRNA^{Ser}$. One must ask then whether a single tRNA could recognize all six serine codons in a primitive eukaryote. This is unlikely since it would involve two groups of codons differing from each other at their first and second positions. More probably, we suggest that the proto-eukaryote already had several serine tRNAs; following gene duplication, a single lineage could then have evolved the various known anticodons and replaced the previous tRNAs which finally became extinct.

3. *The $tRNA^{Val}_{GAC}$-$tRNA^{Gly}$ Superfamily*

Glycine is incorporated into protein in response to four codons of the GGX type. In both eukaryotes and prokaryotes, three isoacceptor families distinguished by their anticodons, UCC, CCC, or GCC, are found. Excluding the *Staphylococcus epidermis* glycine tRNAs involved in cell wall synthesis and the eukaryotic $tRNA^{Gly}_{UCC}$ and $tRNA^{Gly}_{CCC}$ for each of which only one quite different sequence is known, analysis of branching order has shown that the GCC subfamily separated from the U/CCC subfamily prior to the divergence of eukaryotes and prokaryotes.[60] The lack of sequence data has made it impossible to determine the chronology of the UCC-CCC subfamily divergence with respect to the eukaryotic-prokaryotic branching (Figure 11). It is clear, however, that all glycine tRNAs have evolved from a common ancestor. Furthermore, the glycine family was found to cluster into a single tree with prokaryotic $tRNA^{Val}_{GAC}$. Contrary to the comparatively short period of time represented by the previous $tRNA^{Ser}$ tree, the Val-Gly split could well represent a part of the early diversification of tRNAs among the various amino acid codes. Incidentally, based on the rather high degree of divergence found in pairs of tRNAs with different acceptor activities, there is little evidence to support the idea of recent changes in tRNA charging specificity, even though the artificial selection of an *E. coli* tRNA amber suppressor $tRNA^{Tyr136}$ which could be further mutated to a $tRNA^{Gln}$ shows that this is not impossible.[137]

If a lesson should be drawn from the three above tRNA families, it is that no common time-related pattern exists that simultaneously fits all tRNA data. Each isoacceptor family and, possibly, each branch of a given family can have its own peculiar evolutionary history. For instance, the choice between a purine or a pyrimidine at the wobble position was settled very early in $tRNA^{Gly}$ but quite recently in eukaryotic $tRNA^{Ser}$. Also, the UGA and CGA anticodon-containing serine tRNAs from yeast result from a very recent gene duplication, while the corresponding event in eukaryotic $tRNA^{Lys}_{U/CUU}$ is clearly more ancient. Moreoever, the basic concept which emerges is that an isoacceptor family, as depicted by a set of sequences from various organisms, is frequently the end product of both speciation and gene duplications. Thus the evolutionary origin of isoacceptor complements for a given amino acid can be expected to vary in different species.

Finally, it should be noted that each of the three isoacceptor families above has a unique ancestor. Thus, the authors' data favor the idea that the functional diversification of tRNAs results from a divergent, rather than from a convergent, process. Whether this is universally true and can be applied to tRNAs with different acceptor activities, will require much more sequence data for evaluation. In theory, it is also possible that tRNAs with the same acceptor activity from different taxa originated on distinct evolutionary lines. Indeed, the eukaryotic $tRNA^{Tyr}$, with a short extra loop, and its prokaryotic counterpart, with a long extra loop, do not converge toward a common root suggesting independent origins for the two branches.[60]

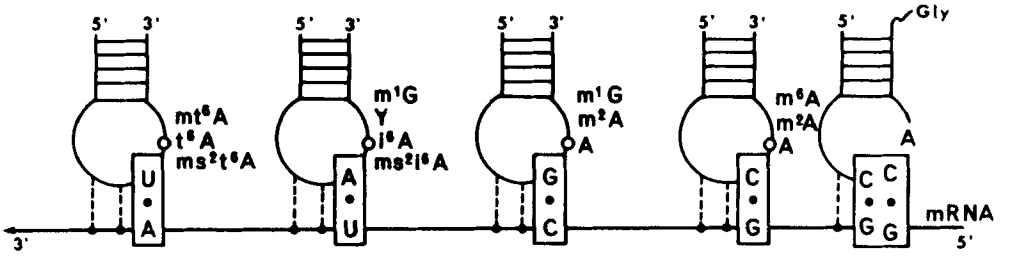


FIGURE 18. Correlation of the anticodon sequence and the 3' adjacent modified nucleoside. A model of the codon-anticodon complex is shown. Abbreviations used are given in Table 5. (From Grosjean, H. and Chantrenne, H., *On Biological Recognition*, Springer-Verlag, Amsterdam, 1979. With permission.)

B. Fine Tuning of Codon-Anticodon Interaction: A Possible Role for Modified Bases

We have previously considered the role of modified nucleosides in tRNA structure with the exception of those found in the anticodon region. This latter group of modifications is often characterized by complex structures which are present in two positions of the tRNA sequence: the 3' adjacent position and the 5' terminal position of the anticodon.

The base adjacent to the 3' side of the anticodon is invariably a purine which is frequently hypermodified. The first indication of a possible role for these modified purines was provided by Nishimura, when he showed a correlation between the anticodon (or codon) structure and the nature of the modification (Figure 18).[138] A remarkable illustration of the effect of these correlations is that when cytosine is replaced by uracil at the third position of the anticodon in a mutant *E. coli* $tRNA^{Gly}$, the normal adenosine next to the anticodon is modified by a posttranscriptional threonylation.[139] In the same tRNA the replacement of the original cytosine by adenine results in the methylthiolation and isopentenylation of the adjacent adenosine.[116] Clearly, the modification is specific not for the tRNA per se, but for a class of anticodons and is therefore related to more general features of the decoding process.

There is now growing evidence that modified nucleosides in the 3'adjacent position are involved in the stabilization of the codon-anticodon interaction.[140] For instance, two immature tRNAs, a suppressor $tRNA^{Tyr}$ having an unmodified A next to the anticodon,[141] and a $tRNA^{Ile}$ lacking t^6A,[142] are unable to bind to ribosomes, although they are normally charged by their cognate synthetases. Similarly, excision of the wye-base from the anticodon loop of $tRNA^{Phe}$ abolishes codon recognition.[143,144] On the other hand, unmodified $tRNA^{Phe}_{GAA}$ from *Mycoplasma*[145] and tRNA from a *Lactobacillus acidophilus* mutant[146] lacking half of its isopentenyladenosine content seem to support protein synthesis in an *E. coli* system, but neither of these two tRNAs have been evaluated for translational accuracy or under competitive conditions with fully modified tRNAs. In addition, mutants, as in *Saccharomyces cerevisiae* whose tRNAs do not contain any isopentenyladenosine, grow perfectly well, although nonsense suppression by tRNA is abolished.[147] This latter observation may be explained by the finding of one coauthor, Henri Grosjean, that the terminator sequence UAG forms an anomalously unstable complex with the complementary tRNA anticodon, suggesting that the stability of the codon-anticodon interaction involving nonsense codons would be more sensitive to the absence of modifications. In the final analysis, the lack of hypermodified nucleosides in mitochondrial tRNAs,[39] indicates that the 3' adjacent modification is not an *absolute* functional requirement, although it could confer some selective advantage to an organism where it is present.

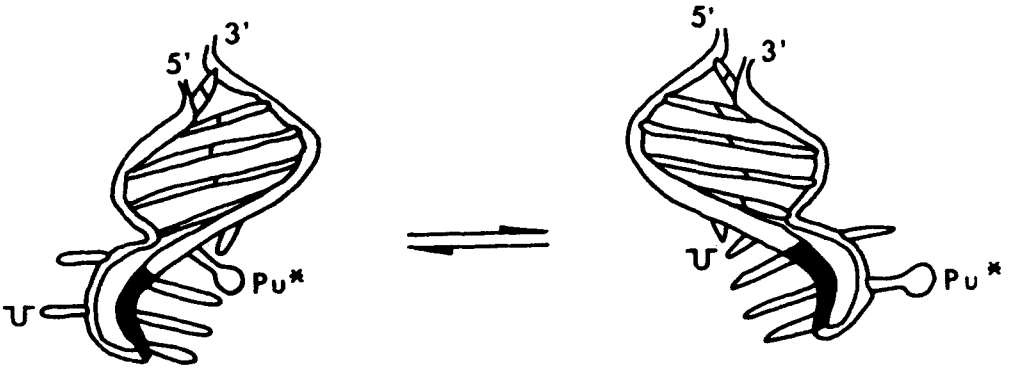


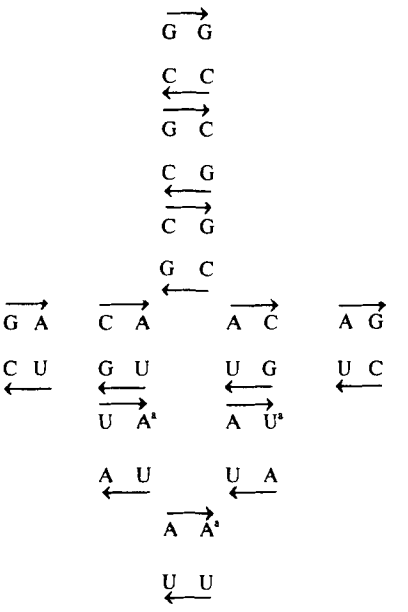
FIGURE 19. The Fuller and Hodgson model of the anticodon loop. The figure on the left is the so-called FH conformation where the 3′ part of the loop is stacked onto the helical anticodon stem. The figure on the right is the hf conformer. In this conformation the nucleosides from the 5′ part of the loop are stacked on the stem. (Adapted from Fuller, W. and Hodgson, A., *Nature (London)*, 215, 817, 1967.)

One may then ask *how* the hypermodification contributes to the overall stabilization of the codon-anticodon complex. First, crystallographic data,[148,149] NMR studies,[150] and fluorescence measurements[151] are all consistent with the idea, originally advanced by Fuller and Hodgson,[152] that the five bases on the 3′ side of the anticodon loop form a 3′ stack continuous with the anticodon stem (Figure 19). Although not favored, these authors could not rule out an alternate 5′ stacked model. In the 3′ stack, the loop is closed by folding back the two remaining nucleotides, thus producing a kink in the backbone at the invariant U_{33} residue (the so-called U-turn).[20] However, the anticodon loop is still quite mobile, both in solution and the crystal.[30] The increase of wye-base fluorescence in the presence of the complementary phenylalanine codon indicates a strong stacking interaction implicating the hypermodified purine in yeast $tRNA^{Phe}$.[153] Similarly, the work of Grosjean has shown that tRNAs form anticodon-anticodon dimer complexes,[154] the stability of which depends on the presence of hypermodification.[155,156] Finally, A or U on the 3′ side of the anticodon is almost always followed by a hypermodified purine; there is no hypermodification, however, if the third (3′) anticodon base can form a strong G-C pair with the codon.[140] The most plausible rationalization of these data is that the codon recognition process implicates the 3′ stacked conformation of the anticodon and that the hypermodifications, characteristic of A and U terminal anticodons, compensate for the notoriously poor stacking of A and U residues[85] (Table 7), thus guaranteeing the continuity and the integrity of the stacked region. The lack of hypermodification in C and G terminal anticodons would compensate for the greater strength of this type of codon-anticodon interaction. Consistent with this view, the half-life of complexes between tRNAs with complementary anticodons is nearly independent of the G-C content of the anticodons.[154]

Fuller and Hodgson also examined the alternate conformer having the five nucleosides on the 5′ end of the loop stacked on the anticodon stem (hf conformation).[152] If, as argued above, the active conformation is 3′ stacked (FH conformation), the invariant U adjacent to the 5′ side of the anticodon can be viewed as a factor destabilizing the rival 5′ stack while providing a pivot for the U-turn. Another corollary of this model is that the first anticodon position (wobble), due to its particular flexibility related to its terminal position in the 3′ stack, should not interact with the same specificity as the second and third position of the anticodon (see Figure 20). Some confirmatory evidence for a normal 3′ stacked conformation can be gleaned from the exceptional case of initiator $tRNA^{Met}$. This tRNA shows ambiguous codon recognition on the 3′ and not the 5′ side of the

Table 7
RELATIVE ORDER OF THE STACKING ENERGY OF BASE PAIRS

Stacking base pairs



Note: The contribution of each combination of the four bases in stacked conformations is shown in decreasing order from the best combination at the top. The 5′ to 3′ direction is shown by the arrow.

[a] Low energy pairs that are always followed by a hypermodified nucleoside when present in the second and third anticodon position.

Adapted from Borer, P., Dengler, B., Tinoco, I., and Uhlenbeck, O., *J. Mol. Biol.*, 86, 843, 1974.

anticodon[157] however, the purine adjacent to the 3′ terminal base of the anticodon is usually not hypermodified and C occasionally replaces the invariant U, 5′ adjacent to the anticodon.[3] Crystallographic[45] as well as nuclease susceptibility[84] data suggest that the anticodon loop of initiator tRNA is indeed different from that of tRNAs involved in elongation.[45]

C. To Wobble or Not, That is the Question

Ambiguity in codon-anticodon interactions which was deduced from the discovery that a single tRNA can respond to several different codons,[33] is generally limited to the third position of the codon during the normal process of protein biosynthesis. With the information available in 1966, Crick summarized observed codon-anticodon interactions with a hypothesis linking the ambiguity to a structural flexibility (wobble) of one of the opposing bases whereby nonstandard hydrogen bonds could be accommodated.[33] The hypothesis was consistent with the demonstration that G in the first position of the anticodon could pair either normally with C or with U by wobbling. The 5′ terminus of the anticodon was thus called the wobble position. Unfortunately, this model has not explained some of the more recent observations regarding codon-anticodon inter-

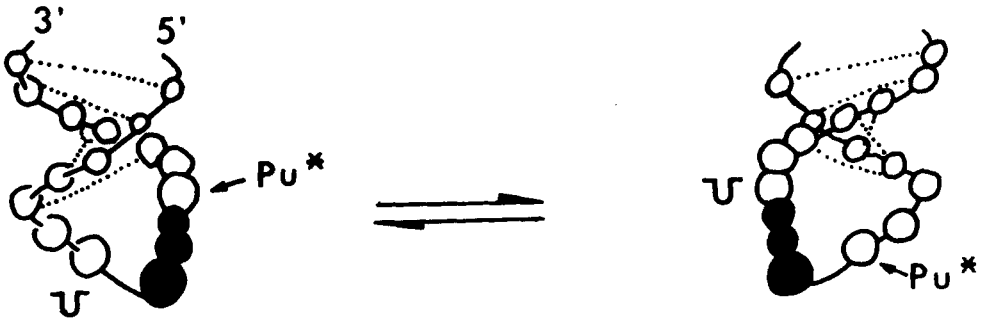


FIGURE 20. Uncertainty in the position of anticodon loop nucleosides. The figure to the left is the FH conformation.[152] The diameter of the balls representing the nucleosides is proportional to the vibrational freedom of each position as determined by X-ray crystallography. On the right is the hf conformation. Both figures show that the position of the last base in the stacked region is much less defined. (Adapted from Kim, S.H., *Transfer RNA*, MIT Press, Cambridge, 1978, 248. With permission.)

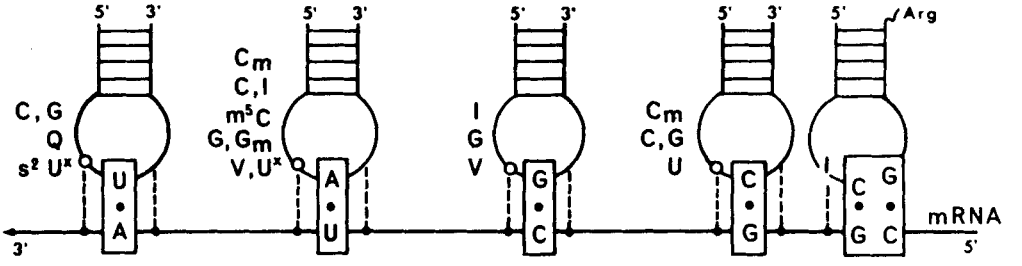


FIGURE 21. Correlation of anticodon sequence with wobble modification. A model of the codon-anticodon interaction is shown. Modified nucleosides are identified by the symbols of Table 5. U* and s^2U* are various derivatives of U and s^2U.

actions.[140,158,159] Although hydrogen bonding may play a role in codon-anticodon complexes, clearly other factors, such as stacking (see above),[155,156] ribosome interactions,[160] kinetic parameters,[161,162] and context effects[163–165] from the mRNA can not be ignored. It now seems appropriate to revise the explanation of decoding ambiguity at the first anticodon position.

Figure 21 summarizes the variety of base modifications found at the wobble position and correlates these with the second position of the anticodon.[107] First, it is noticed that, excluding the two, nonmitochondrial tRNAs where U is the first anticodon base (tRNAs which have G-C pairs in the second anticodon position), the poor stackers, A and U, are never found in the wobble position.[3,85] Only modifications of these bases are permitted, namely inosine instead of adenosine and a variety of modified Us instead of U. Depending on the modification, the ambiguity of the wobble position will be either increased or decreased: for instance, s^2U recognizes only A,[166] while V can read codons ending in U, G, or A.[167]

Except for possible steric hindrance,[154,168] it is our view that the type of interaction allowed at the wobble position will depend essentially on the overall free energy of the tRNA-messenger-ribosome complex which in turn depends on the stacking energy involved. This reasoning explains the puzzling observation that some tRNAs exhibit total ambiguity at the wobble position, that is reading all four codons with identical first and second nucleotides,[169] this shows that the wobble interaction is more a question of conformation and spatial consideration than hydrogen bonding.

It should be noted that Ninio was instrumental in stressing the nature of codon-anticodon interactions based on stacking energy.[170] Logical extension of the data available at that time led him to the stunning "missing triplet hypothesis", since triplets which could compete too successfully with noncognate codons would be dangerous and have to be eliminated. The authors feel that the ability of modified nucleosides to enhance greatly the codon-anticodon interactions as presented above, largely obviates the necessity of this hypothesis.[154]

Incidentally, the model proposed here could easily be completed using the ratchet hypothesis of Woese,[171] since the 3′ stacked conformation necessary in the ribosomal A site codon-anticodon complex could give way to a 5′ stacked conformation in the P site.[172] This possibility has the attractive quality of predicting the decoding properties of initiator tRNAs.[157] If this tRNA is placed in the P site with a 5′ stacked conformation, the ambiguous wobble position then becomes the third base of the anticodon, corresponding to the known ability of this anticodon to interact with either the usual AUG or the unusual GUG codon.[157]

It is important to make one other point bearing on this conformational model, namely that the fidelity of codon-anticodon interaction is not solely dependent on loop bases. First, it is known that the $tRNA^{Trp}$ of *E. coli* which reads the UGA terminator codon is structurally altered in the D region.[118] Second, the interaction between a tRNA anticodon and its complementary codon can trigger conformational changes elsewhere in the tRNA molecule.[35,172–175] Finally, the conformation of $tRNA^{Met}$ and $tRNA^{fMet}$ anticodon loops are distinguishable by a partial S_1 nuclease probe even though their anticodon loop sequences may be identical.[84]

This peculiarity of initiator tRNAs has been correlated to the presence of three G-C (all Gs on 5′ side) base pairs adjacent to the anticodon loop. The effect of these pairs may be twofold: (1) to help stabilize a 5′ stacked conformation in the anticodon loop[84] and (2) to make the transfer from either stacked form to the other more difficult since the stalk will not open as easily as in other tRNAs. The invariance of the anticodon loop-proximal stem region with the $tRNA^{Lys}$, $tRNA^{Phe}$, and eukaryotic $tRNA^{Ser}$ families (Figure 15) could indicate that this part of the molecule modulates the codon-anticodon interaction of several tRNAs. The evidence for this modulation is somewhat ambiguous since this region can be implicated in the synthetase recognition process as well.[96]

D. Relationships to the Evolution of the Genetic Code

Many of the speculations on the origin of the genetic code have centered on the code itself with little regard to the fact that even a primitive code must have required at one point or another some kind of adaptor to decipher it. Taking the alternate viewpoint, the authors wonder if aspects of tRNA structure do not suggest an origin for the code.

The first major consideration is that the anticodon loop of tRNA has determined the triplet nature of the genetic code,[155] although ribosome factors may determine the initial reading frame.[176] The triplet structure may have been selected originally because it permits the formation of codon-anticodon complexes of sufficient energy to engage in protein synthesis. A striking example of the loop effect involves a frameshift suppressor $tRNA^{Gly}$ from *E. coli* which has an extra, eighth, base in the anticodon loop; this tRNA reads the quadruplet GGGG codon.[42] The triplet nature of the code and the architecture of the anticodon loop are, thus, intimately linked.[140]

Going a step further, one may ask why the code is highly degenerate at the third (wobble) position of the codon. Again, it is the authors' feeling that this ambiguity was imposed by the structure of the anticodon loop. In other words, as the 3′ stacked conformation became a permanent feature of protein synthesis, the flexible, helix-terminal nucleotide naturally engaged in less precise interactions.[152] In fact, much in vitro

protein synthesis evidence shows that under some conditions the first anticodon position may have no discriminatory function.[169,176-180] Even in vivo, extreme starvation for an amino acid during protein synthesis leads to a so-called "stuttering" whereby misreading of pyrimidines for purines at the third codon position permits the incorporation of "incorrect" amino acids into protein.[181,182]

Thus, under certain conditions, specificity in codon-anticodon interaction is determined by two of the three bases in the anticodon (the second and third in the anticodon and the first and second in the codon). This "two out of three recognition" can be supported by not only the above experimental data, but also by the basic structure of the genetic code.[183] The topology of the $tRNA^{Gly}$ family of Figure 11 strongly suggests that two out of three recognition is also related to the evolution of the genetic code from simpler, more redundant codes.[60] It remains to be seen, however, if the rules for tRNA isoacceptor divergence deduced from the glycine family are generally applicable.

From the model developed above, it seems only logical to assume that the preceding, two out of three code was converted to the present code by the addition of tRNA isoacceptors and the introduction of modified nucleosides into the wobble position. The lack of these modified nucleosides should have the following consequences:

1. Any base would be permitted in the wobble position
2. Fewer than 20 amino acids could be encoded in the genetic code without increasing coding ambiguity
3. The codon-anticodon interaction would be energetically less stable
4. Protein synthesis would be jerky due to some strong interactions when G or C was present in the wobble position.

One could imagine that an even simpler system could have preceded the above two out of three state by removing the possibility of modifications at the 3′ adjacent nucleoside. From the above arguments the lack of 3′ modifications should favor the 5′ stacked, hf conformation. The actual conformation adopted by the anticodon loop interacting with the codon would then depend on which conformation (5′ stacked or 3′ stacked) permitted the most stable structures. We call this "the best two out of three hypothesis", since geometry of the complex would be determined by the two best base stacking pairs (anticodon 1,2 to codon 2,3 or anticodon 2,3 to codon 1,2). Although this model superficially resembles that of Jukes,[184] in some detail,[186] it is conceptually different since it is based on conformational and thermodynamic arguments. Some experimental work would tend to support a best two out of three concept, namely the $tRNA^{fMet}$ recognition of GUG and AUG codons[157] and the $tRNA^{Cys}$ recognition of arginine codons.[185] Again one could predict that the energy requirement for codon-anticodon interactions at this stage in evolution would be less compared to the preceding two out of three model.

Finally, two other fine adjustments of the codon-anticodon evolving model must be made:

1. There is a strong tendency, as has been noted by the authors and others, that the positions adjacent to codons on the mRNA are not filled completely at random.[163-165,186,187] This nonrandomness is probably responsible for what has been called a "context" effect in the suppression efficiency of nonsense codons.[163-165]
2. There is more and more evidence that codon use is nonrandom[186-189] since, like other translational factors, it should tend toward an optimal efficiency under evolutionary pressure.

Figure 22 is an hypothetical scheme of the evolution of the codon-anticodon interaction based on the above discussion. While it is clear that Step 1 showing the best

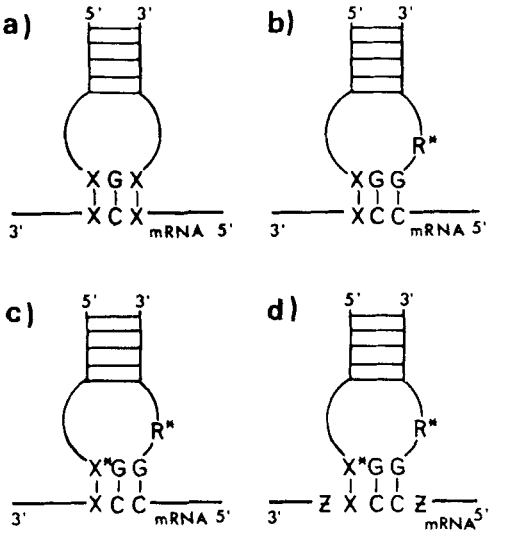


FIGURE 22. Evolutionary model of codon-anticodon interaction. (a) Step 1. Best two out of three model predicts that due to the lack of a modified purine in the 3′ anticodon adjacent position, many tRNAs can exist in either a 3′ or a 5′ stacked conformation. The 5′ stacked conformation permits noncomplementary interactions in the 3′ terminal anticodon position, represented by X-X (where X is any of the four bases), whereas the 3′ stacked conformation permits noncomplementary X-X interactions in the 5′ terminal anticodon position. The conformation selected depends on the relative stability of the alternate forms. (b) Step 2. Classical two out of three recognition, following the introduction of the 3′ adjacent modification. A* is a modified purine when the identity of the third anticodon position is either A or U. (c) Step 3. Modifications introduced at the wobble position restrict the interactions with the third position of the codon. X* is a modified nucleoside in the wobble position. (d) Step 4. Context effect from mRNA and nonrandom use of codons. Z indicates favored bases in these positions which depend on the identity of bases in the anticodon and codon.

two out of three model must, for a given codon family, precede Step 2 (the two out of three model), there is no particular reason why these two models could not coexist in different regions of the genetic code. Steps 3 and 4 involve more the response of mRNA to evolutionary pressure. One aspect, the nonrandom distribution of nucleosides adjacent to the codon, subtly suggests that codon-anticodon interactions involve energetic elements emanating outside of the codon itself. The Step 4 factor of code word choice is thought to be driven by a desire to attain an optimal energy in codon-anticodon interactions by eliminating interactions which are either too stable or not stable enough.[186] Recent work on DNA and RNA sequences has provided other explanations as well: to modulate the rate of protein synthesis,[190] and to avoid ambiguity in mitochondrial protein translation (missing triplets?).[192,193]

	U	C	A	G	
U	Phe 11, 15	Ser 20, 19, 5–NGA, 6	Tyr 6, 11	Cys 2, 4	U
	Leu 6–NAA, 9		Ochre 5	Opal 2	C
			Amber 0	Trp 3	A
					G
C	Leu 4, 5, 3, 79	Pro 0, 8, 5–NGG, 30	His 3, 8	Arg 21, 29, 4, 2	U
			Gln 22–NUG, 34		C
					A
					G
A	Ile 27, 34, 1	Thr 23, 29, 4–NGU, 9	Asn 10, 27	Ser 6, 13	U
	Met 32+12		Lys 75, 12	Arg 2–NCU, 0	C
					A
					G
G	Val 53, 17, 33, 21+1	Ala 63, 25, 41, 41	Asp 24, 32	Gly 34, 37, 1–NCC, 4	U
			Glu 59, 15		C
					A
					G

FIGURE 23. Codon use in *Escherichia coli* compared to T_4-encoded tRNAs. Codon frequency in *E. coli* is derived from known messenger sequences.[189] The T_4 tRNAs are indicated by their anticodon sequence adjacent to the codons which they read.

VII. TWO SPECIAL SYSTEMS: T_4- AND MITOCHONDRIAL-CODED tRNAS

Quite recently, detailed information on tRNAs of two biological systems have emerged which have direct bearing on this review. The tRNAs of bacteriophage T_4 have been extensively studied by McClain, Abelson, Guthrie and their associates, whereas the mitochondrial story has been elaborated by several laboratories in the U.S. and Europe. In both bases, tRNA sequence data and extensive gene mapping have been obtained. We present here a detailed account and evaluation of the evolutionary impact of some of these data.

A. T_4 Bacteriophage tRNA

1. Why T_4 tRNA Genes?

Shortly after infection of host *E. coli* by the bacteriophage T_4 there is an appearance of eight new tRNA species which are encoded by the phage genome.[194] A uridine derivative, N, of unknown structure is present in the first anticodon position of seven of these tRNAs; the exception, $tRNA^{Ile}$, contains a cytidine.[3,107] It is thought that these tRNAs recognize A in the third codon position, although this has not been proved in all cases.[194a] In addition, T_4 directs the synthesis of an enzyme which specifically nicks the host $tRNA^{Leu}$ reading CUX-type codons,[195] the T_4 $tRNA^{Leu}_{NAA}$ would read at least one of the other leucine codons.[196] Although this new set of tRNAs is not necessary for the expression of T_4 genes in all *E. coli* strains, its presence is thought to give the invading genome some advantage in the competition between host and phage gene expression.[196a,196b,197] As a partial confirmation of this hypothesis, consider Figure 23 which compares the presumed anticodon specificities of T_4-encoded tRNA with codon use by the *E. coli* genome. Obviously, the codon use data from *E. coli* are rather limited.[189] A

Table 8
HOST AND PHAGE-ENCODED tRNA COMPARISONS

T_4-Coded tRNAS

tRNA	Anticodon	Number of differences with *Escherichia coli* tRNA
Arg	NCU	39[a]
Gln	NUG	23[a]
Gly	NCC	12
Ile	CAU	28[a]
Leu	NAA	24[a]
Pro	NGG	—
Ser	NGA	19
Thr	NGU	16[a]

Note: N is identified as a modified U. Sequence differences calculated as before.[60]

[a] Comparison done with host tRNA of unlike anticodon.

striking correlation can be made however: T_4 tRNAs read degenerate codons which, as a whole, are underused by *E. coli*. Since the frequency of codon use is generally reflected in tRNA levels,[198,198a] the T_4 tRNAs may then be required to complement *E. coli* tRNAs for these eight codons. Although the codon usage of T_4 genes is not presently known, the authors feel that a nonrandom distribution favoring codons translatable by T_4 tRNAs will be found. Consistent with host codon use, bacteriophages may be thought to carry tRNA genes in order to improve the efficiency of the host translation machinery. Interestingly, in a compilation from eukaryotes and their respective viruses, it can be shown that virus-coded genes generally use codons which are underrepresented in their host genomes.[189] Thus, these results could indicate a type of general biological barrier, based on differences in the translational process between viruses and their hosts.

2. Where Do T_4 tRNAs Come From?

Although controversial, the authors would like to consider the origin of phages with reference to the known tRNA structures. From Table 8, it is abundantly clear that phage tRNAs are quite different from their host counterparts and that the extent of divergence, even between tRNAs with the same anticodon, is quite large. This statement must necessarily be tempered by the realization that all host tRNAs have not as yet been sequenced and, therefore, it is possible that another host isoacceptor may be closer to the phage tRNAs. Only the $tRNA^{Gly}_{UCC}$ and the $tRNA^{Thr}_{UGU}$ show really significant similarities with *E. coli* tRNAs. Given the probable evolutionary conservatism of $tRNA^{Gly}$ in prokaryotes,[60] even this close resemblance indicates only a very distant common ancestor. The lack of similarity between host and phage isocoding tRNAs is indeed unexpected, since frequent genetic exchange between host and phage tRNA genes might have been expected. Moreover, even if phage-coded tRNAs have an independent origin, the sequences, phage and host, should approach each other since they are subjected to identical selective pressure by the host-protein translation machinery.

Because of the little resemblance between host and phage tRNAs, the authors have looked for similarities among phage tRNAs themselves. This experiment is also dictated

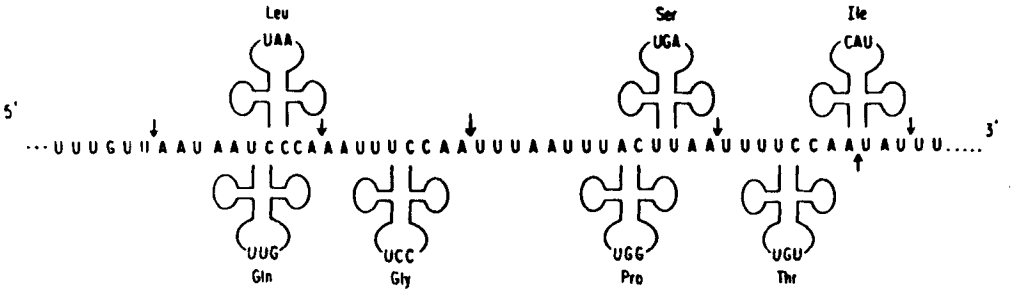


FIGURE 24. Physical map of the T_4-tRNA genetic cluster. The nucleotide sequences intervening between the seven adjacent tRNAs are shown. Arrows indicate positions of early cleavage during the maturation process. The eighth $tRNA^{Arg}$ is located on the 5' side relative to the cluster. (From Abelson, J., *Annu. Rev. Biochem.*, 48, 1035, 1979. With permission.)

by the gene mapping of these phage tRNAs shown in Figure 24;[199] the dimeric structure of some genes is suggestive of divergence after gene duplication. Again, surprisingly, these tRNAs differ greatly in structure; sequence comparisons give little clue to their origin. Perhaps of some significance is that among the comparisons giving the lowest number of differences is that between $tRNA^{Gln}$ and $tRNA^{Leu}$, the products of closely linked genes.

Recently, McClain has discovered an interesting mechanism which may be involved in T_4 tRNA evolution.[194a] Using a T_4 mutant containing only two tRNA genes, he has been able to isolate a hybrid tRNA which apparently results from a recombination event between the two tRNA genes (see Figure 25). Due to the relatively high frequency of these events, he has postulated that this type of gene recombination is an important mechanism in tRNA evolution. Although it may indeed be the case for some T_4 tRNAs (especially $tRNA^{Glu}$ and $tRNA^{Leu}$), there is no compelling evidence that the same process also occurs with cellular tRNA genes. Rather, the similar species branching order in phylogenetic trees for various tRNA families and the predominance of evolution by point mutation deduced from mutation maps would tend to argue against a rapid, noncontinuous evolutionary event as suggested by tRNA-tRNA recombination. However, such a mechanism could be the basis of an editing system proposed to explain the lack of heterogeneity of repetitive genes.[200] Also the apparent shuffling of eukaryotic isoacceptor anticodons alluded to in Section VI.A. may well result from a similar recombination event.

3. The CCA Terminus

By isolation of precursor tRNA dimers, it was also shown that the CCA terminal sequence of phage tRNAs is not encoded in the primary transcript, but must be added during processing by the host ATP(CTP):tRNA nucleotidyl transferase.[201] In addition, eukaryotic and mitochondrial tRNA genes do not contain the CCA terminus sequence,[202,202a] thereby adding strong support to the hypothesis of McClain and coworkers that early tRNA genes did not contain this sequence. The *E. coli* (presumably other prokaryotes as well) tRNA genes do, in fact, code for the CCA terminus[203,204] must then be explained by the acquisition of the trinucleotide sequence by each prokaryotic tRNA gene during its emergence from the ancestor common to these organisms. Reasoning that prokaryotic systems have diverged the most from the common ancestor of different cell types, although not generally considered, has been persuasively presented by Reanney[205] (see also below).

Paradoxically, then, the necessity for the CCA terminus being illustrated by its presence in all mature tRNAs is contrasted with the deduction that the sequence was not

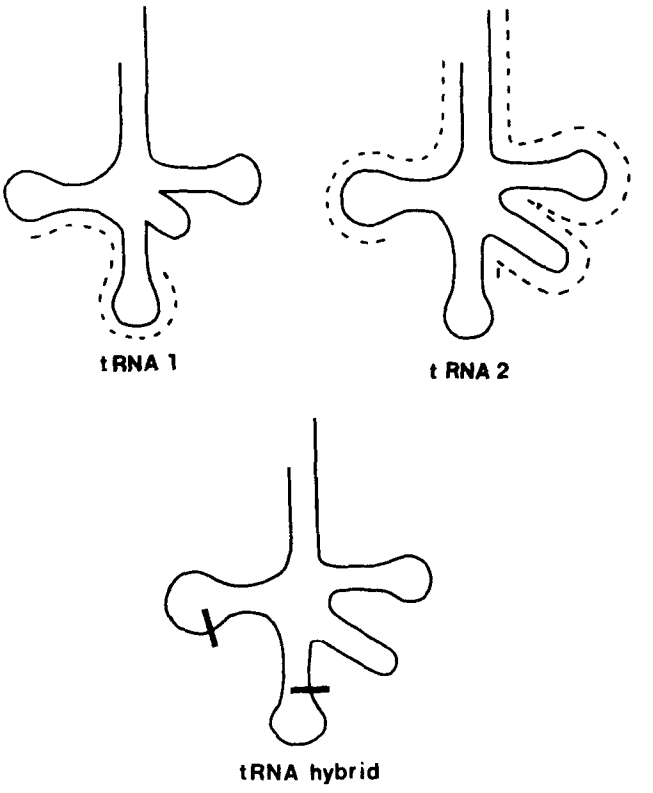


FIGURE 25. Recombination of T_4 tRNA genes. McClain model of tRNA recombination found in T_4 bacteriophage system.[194a]

present in the primordial gene. Perhaps both ideas can be reconciled by the recent suggestion of Dillon that primordial tRNA was nothing other than the CCA terminus.[206] As the translation process became more like the present-day version, the CCA sequence could have been put onto a piece of RNA. In this sense, the maturation of tRNA by the addition of CCA can also be thought of as the polynucleotidylation of the biologically active CCA molecule.

B. Mitochondrial tRNA

1. The Dilemma of Mitochondrial Origin

It is now quite clear, based on tRNA[58] or rRNA[207,208] sequence data and on similarities of the photosynthetic processes, that the chloroplast originated as an endosymbiot of an eukaryotic cytoplasm and shared a common ancestor with modern blue-green algae. On the other hand, with the near revolution in mitochondrial data generated during recent years, the debate rages as to whether the origin of the mitochondria is endosymbiotic or a result of the specialization or compartmentalization of a partial eukaryotic genome.[209-211] In contrast to the size homogeneity of chloroplastic DNA, mitochondrial DNA size varies from 10^7 to 5×10^7 daltons.[212] Mitochondrial functions, however, are highly conserved.[211] Similarly, the number of mitochondrial tRNA genes remains relatively stable at 20 to 25 across major eukaryotic phyla.[213] Based mainly on constant functionality, it is difficult to conceive of anything other than a common or monophyletic origin of mitochondria. However, any general evolutionary model of their origin must take their diversity into account.

Recent work by Bernardi may shed some light on this subject. Using nuclease

digestion, a model of yeast mitochondrial (mt)DNA arrangement has been proposed[214] (cf. Reference 215). MtDNA, which has an extremely high A-T content, is seen as a series of "genetic units" composed each of a G-C rich cluster, a restriction nuclease cluster, a gene and an A-T rich spacer, not necessarily in that order.[214] Differences in DNA length in various "petite" yeast mutant strains are thought to arise by any number of unequal crossing-over events between the homologous A-T rich regions.[216] This model explains the observation that mtDNA from *Saccharomyces cerevisiae* and *S. carlsbergensis* have very different nuclease digestion patterns, while exhibiting very high homology as determined by DNA-DNA hybridization.[217] Extension of this notion of frequent unequal crossing-over may go a long way to explain mtDNA size variation in different eukaryotic phyla which, in turn, would suggest that mtDNA variation is mainly in noncoding regions. The number of mt genes, whether for RNAs or proteins, would remain relatively constant, thus preserving the functional role of mitochondria.

After considering the mt gene structure of unique and repetitive regions, Bernardi has underlined its basic eukaryotic nature in accord with the view that mitochondria have evolved from nuclear type genes rather than from prokaryotic genes.[214] The existence of intervening sequences in at least mt rRNA genes would superficially support this idea.[218] Whereas it is probable that mitochondria did evolve from an ancestor that possessed the above traits, it is not at all clear that this ancestor was an eukaryote.[205,211,219] An equally probable scenario could be put forth that the proto-organism discussed in the above sections had these properties as well.[205]

2. Mt-tRNA Structures

A major "tour de force" of structural biochemistry was the recent exploitation of the mitochondrial system. There are presently several known sequences from yeast, *Neurospora crassa*, and human mitochondrial tRNAs.[39,202,220-227] A quick glance at some of these reveals an extremely high A-U content unparalleled in any other type of tRNAs. As far as sequence comparisons are concerned, mt tRNAs diverge very much (usually 30 differences or more) from their eukaryotic or prokaryotic counterparts, although they are marginally closer to that latter group. The convergence test described in Section II shows no more similarities between mitochondrial sequences and proto-eukaryotic or proto-prokaryotic tRNA than exists between present-day sequences. Due to the present paucity of data, this is hardly conclusive and neither proves nor disproves the endosymbiotic theory of origin.

The mutational rate of mt tRNAs is another unknown, since only 2 out of the 10 sequences correspond to isocoding tRNA. In this case, it is found that yeast and *N. crassa* mitochondrial initiator $tRNA^{Met}$ differ at 26 positions while the cytoplasmic initiators from the same organisms show only 16 differences.[60] Thus, it would seem at least superficially that mitochondrial tRNAs are much more variable than their cytoplasmic counterparts. This single example, however, is not entirely convincing since other cytoplasmic isocoding tRNAs (for instance, yeast and *Schizosaccharomyces pombe* $tRNA^{Phe}$) may sometimes exhibit as much divergence as the two mitochondrial initiators.[60] Clearly, the only way to be able to use mt tRNA structural data for phylogeny or determination of the mutation rate is to have a large sequence set composed of tRNAs having identical anticodons and isolated from all major eukaryotic phyla. Even then, it may be a difficult task because of the almost binary (A-U) composition of these tRNAs.

Of at least equal importance as the gross sequence similarity or dissimilarity of mt tRNA are the unusual characters in what had previously been thought to be functionally important and structurally invariant regions of the tRNA molecule. Some of these curios are

Table 9
OCCURRENCE OF MODIFIED NUCLEOSIDES IN mt tRNA

Modified nucleotide	Number of occurrences	Position[a]
D/D	4/4	16/20
m_2^2G	3	26
Ψ	1	27
Ψ	1	32
Ψ	1	40
Ψ	7	55
m^1G	2	37
G[b]	1	37
A[b]	1	37
m^5C	1	48
m^5C	1	49
T	4	54

Note: Based on seven mt tRNA sequences from *Neurospora crassa* and *Saccharomyces cerevisiae.*

[a] Position number refers to the Cold Spring Harbor Convention. The identity of modified nucleosides is as shown in Table 5.
[b] Unknown modification.

1. In *N. crassa* mt $tRNA^{fMet}$, the TΨC purine (prokaryotic) or AUCG (eukaryotic) sequence of the T loop is replaced by UGCA. The mt tRNA does contain a standard base pair at the end of the acceptor arm as in eukaryotes; these two nucleotides are unpaired in prokaryotes. Finally, the "invariant" GG sequence of the D loop is missing in this mt tRNA.[220]
2. In yeast mt $tRNA^{Phe}$, an extra nucleotide is present in the stem of the T arm.[39]
3. In *N. crassa* mt $tRNA^{Tyr}$, the "invariant" A and purine in positions 14 and 15, respectively, are absent.[221]
4. In yeast mt $tRNA^{Ser}_{GCU}$, there is an extra nucleotide between the dihydrouracil and the anticodon stems; also, the "invariable" purine in position 15 is absent.[224] This last feature is found in yeast mt $tRNA^{Cys}$ as well.[202]
5. Mt $tRNA^{Gly}_{UCC}$ contains three G-U pairs, one each at the base of the D, acceptor, and and anticodon stems.[224]
6. The two human sequences of $tRNA^{Asp}$ and $tRNA^{Lys}$ have abbreviated D loops which, in addition, do not contain the "invariant" GG sequence.[226]

In all, one is left with the impression that, rather than exhibiting a new set of structure-function rules analogous to those of eukaryotic or prokaryotic sequences, mt tRNA may simply have less rules.

Another facet of mt tRNA structure is the paucity of modified nucleosides as indicated in Table 9. Most notable is the lack of hypermodified nucleosides in the wobble position or in the position 3′ adjacent to the anticodon.[39] Unmodified U in the wobble position confers strict two out of three reading to the tRNA. From arguments developed in Section VI., it would not be surprising that the stabilization energy of the codon-anticodon interaction is less than in eukaryotes and prokaryotes. In this sense a possible analogy can be made between the low level of modified nucleosides in mitochondrial tRNA and the comparable situation in *Mycoplasma.*[145,228]

3. The Mitochondrion as an Historical Relic

Major findings have recently been reported concerning anomalies in the mt protein translation system. First, UGA, a termination signal in other systems, seems always to be read as a tryptophan codon by yeast mt $tRNA^{Trp}$.[229] This situation which has been confirmed in human[226] and *N. Crassa*[227] mitochondria is similar to the UGA suppressor $tRNA^{Trp}$ which has been isolated from *E. coli.*[118] A further discovery is that one of the yeast mt $tRNA^{Thr}$ inserts a threonine in response to a CUA "leucine" codon.[222] Finally, AUA normally coding for isoleucine has been taken over by methionine in human mitochondria.[226] Before assuming that these facts can be related to a primitive genetic code, it is worthwhile to consider the general pattern of anomalies in tRNA structure which could also be indicative of a generally less precise translation process.

At least, in some of the above codon equivalents, it is clear that corresponding anamolies are not present in all mitochondrial systems. One, therefore, must not only get used to the idea of a nonuniversal code, but also the probability that the code is not even constant among mitochondria. Assuming that mitochondrial origin is monophyletic (no longer as certain as it might have seemed), the genetic code must have changed since the emergence of the modern eukaryote. Rather than being related to the identity of the primitive code, these data may be related to the mechanism of the establishment of the genetic code. Thus the apparent commandeering of leucine codons by threonine, in the yeast line, is suggestive of processes that may be responsible for the scattering of those triplets which code for Ser and Arg in the present-day code.

Although the interpretation of the genetic code-related results may be premature, other hypotheses on mt translation may now be pertinent. The first obvious point is that the mt tRNA complement (20 to 25) is insufficient to decode 62 codons using the standard "wobble" rules. Therefore, in order to entertain proper protein translation, one or more of the following options are open: (1) tRNAs decode more codons than usual in cellular systems, (2) not all codons are actually used, or (3) cytoplasmic tRNAs are imported.

The first hypothesis is clearly operative: $tRNA^{Trp}$ reads two codons.[229] In addition, the lack of modification in the "wobble" position, modifications which frequently limit wobbling, would open the door to additional interactions for a given anticodon. New data from *N. crassa* convincingly support the expanded reading ability of anticodons containing unmodified uridine in the wobble position.[227] The second hypothesis operates as well: comparison of the ATPase subunit 9 sequence with that of its gene shows an extremely skewed code-word usage with a very strong bias in favor of codons ending in A or U.[192,193] Although some code words are not used at all (see Figure 26), this could be the result of the limited sampling currently available.

Finally, only a small amount of data support the model of tRNA importation. In the case of yeast mitochondria, only the $tRNA^{Lys}$ seems to be imported.[230] However, *Tetrahymena* may be an exception (see below); here, present information favors the importation model for many tRNAs.[231] If, as suggested above, a peculiar pattern of codon-anticodon interactions exists in mitochondria, the introduction of large amounts of cytoplasmic tRNAs could be cataclysmic for mitochondrial translation since these tRNAs could bind more strongly to the codon-ribosome complex and could misread some of the mitochondrial codons as previously suggested.

The biological consequence seems to be particularly crucial for the interpretation of mitochondrial data; based on the above differences in translation machinery, it is likely that a barrier, as in the T_4 system, exists between cytoplasmic and mitochondrial genetic information. The slightly different genetic code in the two systems and the unusual mt tRNA complement may be determinant forces in the conservation of an independent mitochondrial translation apparatus.[232] This barrier will make it difficult, if not

1 \ 2	T	C	A	G	3
T	Phe 1 6	Ser 0 0 5 0	Tyr 1 0	Cys 1 0	T
	Leu 11 0		Term 1 0	Term 0	C
				Trp 0	A
					G
C	Leu 0 0 1 0	Pro 1 0 1 0	His 0 0	Arg 0 0 0 0	T
			Gln 1 0		C
					A
					G
A	Ile 7 2 0	Thr 0 0 2 0	Asn 1 1	Ser 0 0	T
	Met 1·2		Lys 2 0	Arg 1 0	C
					A
					G
G	Val 1 0 5 0	Ala 5 1 4 0	Asp 0 1	Gly 8 0 2 0	T
			Glu 1 0		C
					A
					G

FIGURE 26. Codeword use in mitochondria. This codeword usage distribution was derived from the sequence of the subunit A of mitochondrial ATPase from yeast. (From Hensgens, L. A. M., Grivell, L. A., Borst, P., and Bos, J. L., *Proc. Natl. Acad. Sci. U.S.A.*, 76, 1663, 1979. With permission.) The UAA leucine codon has been shown to direct the incorporation of threonine.[222]

impossible, for mitochondrial mRNAs to be translated on cytoplasmic ribosomes and vice versa. The one current exception to this barrier is the case of *Tetrahymena* where, presumably, large amounts of cytoplasmic tRNA are imported.[231] Evaluation of the possibility that the high A-T content[233] of *Tetrahymena* DNA as in mitochondrial DNA reduces the translational barrier because of the convergent responses to the translation problem will have to await more detailed work on this organism.

Given all the previous considerations, it seems very unlikely that mitochondria could have originated from the compartmentalization of a proto-eukaryotic cell, thereby establishing two coexisting translation apparatuses. The mitochondrion does not seem to be of recent prokaryotic origin either. Although the formylation of initiator $tRNA^{Met}$ and the large size of the extra loop of $tRNA^{Tyr}$ characterize both mitochondrial and prokaryotic systems, they could well be remnants of a more primitive organism. The extensive divergence between prokaryotic and mitochondrial tRNAs would indeed indicate that the two lines separated very early, quite possibly before the eukaryotic-prokaryotic split. To the delight of evolutionists, characteristics of the mitochondrial translation system indicate that mitochondria may, in fact, represent a survivor of an archaic cell type.[211]

VIII. CONCLUSIONS AND OUTLOOK

In reviewing the information relative to the evolution of tRNA structure and function, the authors were struck by the frequent lack of evolutionary interpretations of tRNA data. On the other hand, most evolutionary theses are seldom related to experimental work. The authors have attempted to bridge this gap at least with respect to the subjects

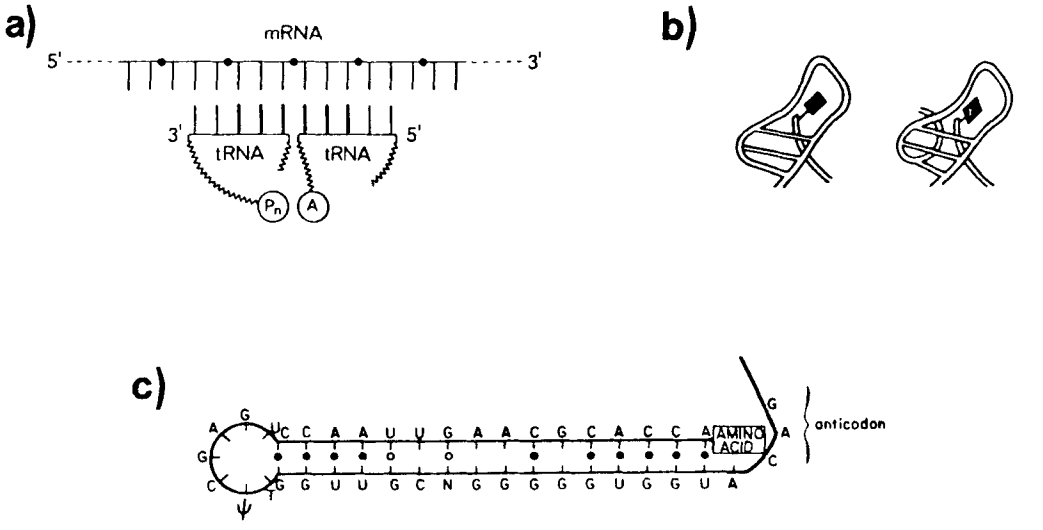


FIGURE 27. Primordial tRNAs. (a) Crick et al. model showing five base-pairs between primitive tRNAs and mRNA. The anticodon loop of the two tRNAs assume the different FH and hf conformations. (From Crick, F.H.C., Brenner, S., Klug, A., and Pieczenik, G., *Origins Life*, 7, 389, 1976. With permission.) (b) Position of amino acid and stem relative to primitive loop in Ninio model. This prototype involves two RNA pieces which are joined later as the molecule is elaborated to its present size and form. The loop shown eventually becomes the D loop of modern tRNA. From Ninio, J., *Approches Moléculaires de l'Evolution*, Masson, Paris, 1979, 83. (c) Hopfield model showing proposed base-pairing scheme which brings the anticodon sequence close to the amino acid acceptor.[161] The portions of the molecule that are not shown would be added during the emergence of the modern ancestor. Relics of the early base-pairing scheme have been found in present-day tRNA structures. From Hopfield, J.J., *Proc. Natl. Acad. Sci. U.S.A.*, 75, 4334, 1978. With permission.)

which have been treated here. In other areas of tRNA biochemistry there are generally not enough experimental results to do a proper evolutionary interpretation. Such is the case of intervening sequences in tRNA and of tRNA-like viral RNAs. Although splicing appears widespread in eukaryotes, there is no evidence at this time that this mechanism was used to generate mosaic tRNAs out of fragments with different origins. Indeed, eukaryotic $tRNA^{Phe}$, although spliced during maturation,[234] has a normal phylogenetic tree, and there seems no doubt that all $tRNA^{Phe}$ sequences have evolved from a single ancestral gene, whether originally split or not.[60]

Another intriguing observation is that some viral RNA sequences can be charged with amino acids by host aminoacyl-tRNA synthetases.[235] This phenomenon, first identified in plant viral tRNAs, now seems to be quite widespread.[20] Even though viral RNA shows only a limited similarity with cellular tRNA, it is apparently charged in vivo and processed to some extent.[236] The evolutionary significance of these tRNA-like structures, however, cannot be evaluated at the present.

Another subject which this review has not discussed until now is the speculation which surrounds the nature of the predecessor of the modern tRNA molecule. First, the common structural elements of present-day tRNAs would suggest that the immediate ancestor of tRNAs resembled very much the modern version. It is inconceivable, however, that such a complex ancestral structure could spring from a void. Several schemes have been advanced to bridge the gap from what must have been a simple beginning to the modern ancestral sequence or sequences. One such hypothesis by Crick et al. involves a primitive adaptor (tRNA) where a seven-base anticodon loop would bind to mRNA five bases at a time in ribosome-less protein synthesis.[237] The actual reading frame, however, is triplet (see Figure 27). A feature of this model is the incorporation of

the Woese ratchet model 3′ stack to 5′ stack-conformational change.[171] The implied sequence overlap would put serious restrictions on the mRNA sequence, but has the advantage of using five base-pairs to stabilize the complex. As pointed out above, however, it is not the number of paired bases which guarantees a stable codon-anticodon interaction, but rather whether a highly stacked complex can be formed.

Two other hypotheses involve a smaller primordial "tRNA". In the Ninio model, the amino acid stem is tucked into the groove of a looped structure (Figure 27b).[238] Initially, these two features are not covalently bound, but as the structure evolves to a higher complexity they are linked, the original loop becoming the D loop of the modern tRNA. This model is rather attractive since the geometry of the original complex is constant throughout the development to the present-day structure. In fact, the hypothesis of Dillon (Section VII.A.) concerning the CCA terminus is quite similar, at least superficially.[206] The Hopfield model, on the other hand, addresses the problem of the physical distance between the two "active" portions of the tRNA molecule: the anticodon loop and the acceptor stem.[161] His model is given some credence, since relics of a predicted base-pairing scheme shown in Figure 27c are still found in modern tRNA.

Eigen and Winkler-Oswatitsch suggest that the primordial tRNA acted not only as an adaptor during translation, but also as a self-replicating messenger.[64] The evidence they derive from sequence comparisons (Section II.I.) suggests that the early sequence consisted basically of 24 in-frame triplets of the form RNY (R: purine, Y: pyrimidine, N: any nucleotide), more specifically GNC. This would provide the high stability of pairing required to achieve replication and translation in a primitive environment. The sequence evidence seems largely to consist of a predominance of Gs at the 5′ end of such "codons" but this is hardly conclusive since the base-paired regions of the tRNA cloverleaf are dominated by G-C pairs. One may predict from this model of repeated, in-frame, triplets that present-day tRNAs could still contain statistically significant repeats. This was indeed found by Ehran et al., both within a tRNA molecule and between tRNAs with different acceptor activities.[239] The "homologous" oligonucleotides were suggested to represent the primitive adaptor which gave rise to the modern tRNA molecule.

Although these models all tend to build larger molecules from smaller ones, it is entirely conceivable that the pre-tRNAs were larger than present-day tRNAs. They could have been pieced together by enzymes involved in RNA splicing if that process existed in primordial times. In fact, it is the authors' feeling that it is worth considering that a prototype tRNA may have served as a message in a crude translation system.[64] Any similarity to the known properties of viral RNAs is not coincidental.

One goal of this review was to organize tRNA structural data into a logical form to deduce evolutionary concepts. From the results presented here, it seems appropriate to make a plea for a more organized approach to the choice of tRNAs to be sequenced, since the quantity of tRNA data that can be generated is obviously finite. As a first goal the authors feel that efforts should be made to establish phylogenetic relationships between as many cell types as possible. This can be most efficiently and effectively accomplished by concentrating on sequences of $tRNA^{Phe}$ or $tRNA^{fMet}$. Among tRNAs presently under study in R. J. Cedergren's, one coauthor's, laboratory are $tRNA^{Phe}$ from methanogens, halophiles and anaerobic bacteria. There are likely to be other types of prokaryotes that may even be more interesting. Moreover, the extreme divergence of *Schizosaccharomyces pombe* $tRNA^{Phe}$ may be just the tip of an iceberg in that the authors expect that many lower eukaryotes will have very unusual tRNA structures. One case in particular, *Tetrahymena* tRNA, is likely to be unique because of the high A-T content of the DNA.[233] Also, the comparative study of mitochondrial tRNAs has just begun. Many more sequences are required to establish a mitochondrial phylogeny, especially if, as is suspected by the authors, mt tRNAs fix mutations more rapidly than their cytoplasmic

counterparts. Although this work may be arduous in mitochondria, the reward could be commensurate: it is possible that a real "fossil" with an archaic genetic code is being dealt with. Finally, a concerted effort will be needed to elucidate the origin of complex isoacceptor families. The number of required sequences will undoubtedly be larger than the number when only phylogenies are sought. The data from this study, as shown in Section VI., could explain the degeneracy of the genetic code. Particularly intriguing are the cases of serine, arginine, and leucine which have six codons each. The lack of similarity between eukaryotic and prokaryotic branches of $tRNA^{Tyr}$ or $tRNA^{Val}$ raises the possibility that tRNAs specific for the same amino acid may have had diverse origins.

In summary, this review has shown that:

1. The fixation of mutations in tRNAs has been slow enough to leave traces of phylogenetic relationships.
2. The determinations of these phylogenetic relationships requires sophisticated statistical methods and a good deal of caution.
3. As expected from the known structural and functional requirements of tRNA, the mutational process has extensively changed some parts of the tRNA molecule while preserving most of the crucial conformational parameters.
4. The triplet nature of the genetic code and its pattern of degeneracy are most easily explained by consideration of the conformational and energetic phenomena related to the interaction between a single-stranded RNA (mRNA) and a RNA loop structure (anticodon loop).
5. T_4- and mitochondrial-coded tRNAs may testify to the existence of translational barriers in complex systems involving the translation of coexisting messages.

ACKNOWLEDGMENTS

The authors of this review were fortunate to have the very real support of many colleagues, who graciously provided unpublished results and helpful discussions. They wish to thank particularly Drs. Stuart Clarkson, James Dahlberg, Guy Dirheimer, Hans Gross, Manfred Eigen, Masami Hasegawa, Gérard Keith, Helga Kersten, Nancy Martin, William McClain, U. L. RajBhandary, Kurt Randerath, Alexander Rich, Dieter Söll, and Gordon Tener for providing material prior to publication, as well as Drs. Jacques Lapointe and Robert Zimmerman for helpful discussions. They would like to express their gratitude as well to Drs. Dieter Söll, William McClain, and Richard Grantham for suggestions, criticisms, and corrections on the first draft manuscript. In addition, the technical help of Nicole Newhouse and Esther Cloutier is appreciated. Original research was supported by the Ministère de l'Education du Québec and the Conseil de Recherches Nationales du Canada. Henri Grosjean received travel support from the Fonds National de la Recherche Scientifique and the Ministère de l'Education Nationale Belge.

NOTE ADDED IN PROOF

During the prepublication period of this review, the number of known mitochondrial tRNA sequences has dramatically increased. The known sequences now include 18 from human mitochondria, 20 from beef, 23 from yeast, 20 from *Aspergillus nidulans* and 9 from *Neurospora crassa*. Many of these sequences were obtained by DNA sequencing techniques.*

* The authors thank Susan Bonitz, Hans Küntzel, Nancy Martin, U.L. RajBhandary and Bruce Roe for communicating to us their unpublished results.

Therefore, the authors thought it useful to update the mitochondrial section of this review with the following results:

1. The hypothesis that a tRNA with an unmodified U in the wobble position can read a four codon box in the genetic code has largely been confirmed.[227,241,242]
2. The fact that yeast mt $tRNA^{Trp}$ can function as a UGA suppressor in in vitro eukaryotic translation systems confirms the use of UGA in mitochondria as a stop codon.[243] In addition, these same experiments show that the ability to read UGA codon is an intrinsic property of the tRNA.
3. A most striking structural anomaly is exhibited by the mammalian mt $tRNA^{Ser}$ which lacks virtually the entire D stem.[244-245]
4. Although some exceptions do occur, isocoding tRNAs from the three fungal mitochondria appear to be evolutionarily related[240] supporting the idea of a monophyletic origin of fungal mitochondria.
5. Isocoding tRNAs from the human placenta and beef liver mitochondria often exhibit an excessively large number of structural differences, although cytoplasmic tRNA structures are almost certainly identical. The divergence of mammalian mt tRNAs may be due to mutational fixation rates as much as two orders of magnitude higher than the nuclear-encoded genes and/or frequent codon specificity changes in tRNAs.

REFERENCES

1. **Dayhoff, M. O. and McLaughlin, P. J.,** Early evolution: transfer RNA, in *Atlas of Protein Sequence and Structure,* Vol. 5, Dayhoff, M. O., Ed., National Biomedical Research Foundation, Washington, D.C., 1972, 111.
2. **Woese, C. R.,** The basic nature of the genetic code, in *The Genetic Code,* Harper and Row, New York, 1967, 150.
3. **Sprinzl, M., Grueter, F., Spelzhaus, A., and Gauss, D. H.,** Compilation of tRNA sequences, *Nucl. Ac. Res.,* 8, r1, 1980.
4. **Monod, J.,** Evolution, in *Le Hasard et la Nécessité,* Editions du Seuil, Paris, 1970, 135.
5. **Zuckerkandl, E. and Pauling, L.,** Paleogenetics, molecular restoration, studies of extinct forms of life, in *Horizons in Biochemistry,* Kasha, M. and Pullman, B., Eds., Academic Press, New York, 1972, 189.
6. **Dobzhansky, T.,** Genetic continuity and change, in *Genetics of the Evolutionary Process,* Columbia University Press, New York, 1970, 30.
7. **Blumenthal, T. and Carmichael, G. G.,** RNA replication, *Annu. Rev. Biochem.,* 48, 525, 1979.
8. **Bishop, M. J.,** Retroviruses, *Annu. Rev. Biochem.,* 47, 35, 1978.
9. **Hanawalt, C. P., Cooper, K. P., Ganesan, A. K., and Smith, A. C.,** DNA repair in bacteria and mammalian cells, *Annu. Rev. Biochem.,* 48, 783, 1979.
10. **Kimura, M.,** Evolutionary rate at the molecular level, *Nature (London),* 217, 624, 1968.
11. **Wilson, A. C., Carlson, S. S., and White, J. T.,** Biochemical evolution, *Annu. Rev. Biochem.,* 46, 573, 1977.
12. **Margoliash, E., Fitch, W. M., and Dickerson, R. E.,** Molecular expression of evolutionary phenomena in the primary and tertiary structures of cytochrome *c,* in *Molecular Evolution: Biochemical Evolution and the Origin of Life,* Schoffeniels, E., Ed., North-Holland/Elsevier, Amsterdam, 1971, 52.
13. **Fitch, W. M. and Margoliash, E.,** Construction of phylogenetic trees, *Science,* 155, 279, 1967.
14. **Sneath, P. H. A. and Sokal, R. R.,** Numerical taxonomy, in *Principles and Practice of Numerical Classification,* W. H. Freeman and Co., San Francisco, 1973.
15. **Weissman, M. S.,** Current approaches to analysis of the nucleotide sequence of DNA, *Anal. Biochem.,* 98, 243, 1979.
16. **Sankoff, D. and Cedergren, R. J.,** A test for nucleotide sequence homology, *J. Mol. Biol.,* 77, 159, 1973.
17. **Cedergren, R. J., Cordeau, R. J., and Robillard, P.,** On the phylogeny of tRNAs, *J. Theor. Biol.,* 37, 209, 1972.
18. **Hatfield, D., Matthews, C. R., and Rice, M.,** Aminocyl-transfer RNA populations in mammalian cells chromatographic profiles and patterns of codon recognition, *Biochim. Biophys. Acta,* 564, 414, 1979.
19. **Holmquist, R., Jukes, T. H., and Pangburn, S.,** Evolution of transfer RNA, *J. Mol. Biol.,* 78, 91, 1973.

20. **Rich, A. and RajBhandary, U. L.,** Transfer RNA: molecular structure, sequence and properties, *Annu. Rev. Biochem.*, 45, 805, 1976.
21. **Roberts, R. J.,** Structures of two glycyl-tRNAs from *Staphylococcus epidermis, Nature (London) New Biol.*, 237, 44, 1972.
22. **Soffer, R. L.,** Aminoacyl-tRNA-protein transferases, in *tRNA: Biological Aspects,* Söll, D., Abelson, J., and Schimmel, P., Eds., Cold Spring Harbor Laboratory, New York, 1980.
23. **Brenchley, J. E. and Williams, L. S.,** Transfer RNA involvement in the regulation of enzyme synthesis, *Annu. Rev. Microbiol.*, 29, 251, 1975.
24. **Sawyer, R. C., Harada, F., and Dahlberg, J. E.,** Virion-associated tRNA primer for Rous Sarcoma Virus DNA synthesis: isolation from uninfected cells, *J. Virol.*, 13, 1302, 1974.
25. **Sigler, P. B.,** An analysis of the structure of tRNA, *Annu. Rev. Biophys. Bioeng.*, 4, 477, 1975.
26. **Kim, S. H.,** Three-dimensional structure of transfer RNA, *Progr. Nucleic Acid Res. Mol. Biol.* 17, 181, 1976.
27. **Goddard, J. P.,** The structure and function of transfer RNA, *Progr. Biophys. Mol. Biol.*, 32, 233, 1977.
28. **Holbrook, S. R., Sussman, J. L., Warrant, R. W., and Kim, S. H.,** Crystal structure of yeast phenylalanine transfer RNA, II. Structural features and functional implications, *J. Mol. Biol.*, 123, 631, 1978.
29. **Clark, B. F. C.,** General features and implications of primary, secondary and tertiary structure, in *Transfer RNA,* Altman, S., Ed., MIT Press, Cambridge, 1978, 14.
30. **Kim, S. H.,** Crystal structure of yeast $tRNA^{Phe}$: its correlation to the solution structure and functional implications, in *Transfer RNA,* Altman, S., Ed., MIT Press, Cambridge, 1978, 248.
31. **Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, J. T., Merril, S. H., Penswick, J. R., and Zamir, A.,** Structure of a ribonucleic acid, *Science,* 147, 1462, 1965.
32. **Sprinzl, M. and Cramer, F.,** Site of aminoacylation of tRNAs from *Escherichia coli* with respect to the 2'- or 3'- hydroxyl group of the terminal adenosine, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 3049, 1975.
33. **Crick, F. H. C.,** Codon-anticodon pairing: the wobble hypothesis, *J. Mol. Biol.*, 19, 548, 1966.
34. **Erdmar, V. A., Sprinzl, M., and Pongs, O.,** The involvement of 55 RNA in the binding of tRNA to ribosomes, *Biochem. Biophys. Res. Commun.*, 54, 942, 1973.
35. **Schwarz, V., Lührmann, R., and Gassen, H. G.,** On the mRNA induces conformational change of AA-tRNA exposing the TΨCG sequence for binding to the 50S ribosomal subunit, *Biochem. Biophys. Res. Commun.*, 56, 807, 1974.
36. **Wrede, P. and Erdmar, V. A.,** *Escherichia coli* 5S RNA binding proteins L18 and L25 interact with 5.8S RNA but not with 5S RNA from yeast ribosomes, *Proc. Natl. Acad. Sci. U. S. A.*, 74, 2706, 1977.
37. **Schulman, L. H., Peska, H., and Sundari, R. M.,** Structural requirements for the recognition of *Escherichia coli* initiator and non-initiator transfer ribonucleic acids by bacterial T factor, *J. Biol. Chem.*, 249, 7102, 1974.
38. **Singer, C. E. and Smith, J. R.,** Histidine regulation in *Salmonella typhimurium,* XIII. Nucleotide sequence of histidine transfer ribonucleic acid, *J. Biol. Chem.*, 247, 2989, 1972.
39. **Martin, R. P., Schneller, J. M., Sibler, A. P., Stahl, A., and Dirheimer, G.,** Yeast mitochondrial tRNAs: isoacceptors, coding origin and primary structure of $tRNA^{Phe}$, in *tRNA: Structure, Properties and Recognition,* Schimmel, P., Söll, D., and Abelson, J., Eds., Cold Spring Harbor Laboratory, New York, 1980.
40. **Wetzel, R., Kohli, J., Altruda, F., and Söll, D.,** Identification and nucleotide sequence of the *sup8-e* UGA-suppressor leucine tRNA from *Schizosaccharomyces pombe, Mol. Gen. Genet.*, 172, 221, 1979.
41. **Ozeki, H.,** Recognition mutants in *E. coli,* in *tRNA: Biological Aspects,* Söll, D., Abelson, J., and Schimmel, P., Eds., Cold Spring Harbor Laboratory, New York, 1980.
42. **Riddle, D. L. and Carbon, J.,** Frameshift suppression: a nucleotide addition in the anticodon of a glycine transfer RNA, *Nature (London) New Biol.*, 242, 230, 1973.
43. **Jank, P., Riesner, D., and Gross, H. J.,** Rabbit liver $tRNA_i^{Val}$. II. Unusual secondary structure of the TΨC stem and loop due to a U_{54}:A_{60} base pair, *Nucl. Ac. Res.*, 4, 2009, 1977.
44. **Schewitz, R. W., Podjarny, A. D., Krishnamachari, N., Hughes, J. L., Sigler, P.B., and Sussman, J. L., Crystal structure of a eukaryotic initiator tRNA,** *Nature (London),* 278, 188, 1979.
45. **Rich, A.,** Personal communication.
46. **Fitch, W. M.,** Aspects of molecular evolution, *Annu. Rev. Genet.*, 7, 343, 1973.
47. **Sankoff, D., Morin, A. M., and Cedergren, R. J.,** The evolution of 5S RNA secondary structures, *Can. J. Biochem.*, 56, 440, 1978.
48. **Sankoff, D.,** Matching sequences under deletion/insertion constraints, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 4, 1972.
49. **Sankoff, D., Morel, C., and Cedergren, R. J.,** Evolution of 5S RNA and the non-randomness of base replacement, *Nature (London) New Biol.*, 245, 232, 1973.
50. **Sankoff, D., Cedergren, R. J., and Lapalme, G.,** Frequency of insertion/deletion, transversion and transition in the evolution of 5S Ribosomal RNA, *J. Mol. Evol.*, 7, 133, 1976.
51. **Fitch, W. M. and Farris, J. S.,** Evolutionary trees with minimum nucleotide replacements from amino acid sequences, *J. Mol. Evol.*, 3, 263, 1974.

52. **Moore, G. W., Barnabas, J., and Goodman, M.,** A method for constructing maximum parsimony ancestral amino acid sequences on a given network, *J. Theor. Biol.,* 38, 459, 1973.
53. **Jukes, T. H. and Holmquist, R.,** Estimation of evolutionary changes in certain homologous polypeptide chains, *J. Mol. Biol.,* 64, 163, 1972.
54. **Moore, G. W., Goodman, M., Callaghan, C., Holmquist, R., and Moise, H.,** Stochastic *versus* augmented maximum parsimony method for estimating superimposed mutations in the divergent evolution of protein sequences. Methods tested on cytochrome *c* amino acid sequences, *J. Mol. Biol.,* 105, 15, 1976.
54a. **Karp, R. M.,** Reducibility among combinatorial problems, in *Complexity of Computer Computations,* Miller, R. E. and Thatcher, J. W., Eds. Plenum Press, New York, 1972, 85.
55. **Sankoff, D. and Rousseau, P.,** Locating the vertices of a Steiner tree in an arbitrary metric space, *Math. Programming,* 9, 240, 1975.
56. **Kimura, M.,** The neutral theory of molecular evolution, *Sci. Am.,* 241, 98, 1979.
57. **Schwartz, R. M. and Dayhoff, M. O.,** Origins of prokaryotes, eukaryotes, mitochondria and chloroplasts, *Science,* 199, 395, 1978.
58. **LaRue, B., Cedergren, R. J., Sankoff, D., and Grosjean, H.,** Evolution of methionine initiator and phenylalanine transfer RNAs, *J. Mol. Evol.,* 14, 287, 1979.
59. **Ninio, J.,** *Approaches Moléculaires de l'Evolution,* Masson, Paris, 1979, chap. V.
60. **Cedergren, R. J., LaRue, B., Sankoff, D., Lapalme, G., and Grosjean, H.,** Convergence and minimal mutation criteria for evaluating early events in tRNA evolution, *Proc. Natl. Acad. Sci. U.S.A.,* 77, 2791, 1980.
61. **Hasegawa, M.,** Evolution of transfer ribonucleic acid, in *Molecular Evolution and Polymorphism,* Kimura, M., Ed., National Institute of Genetics, Japan, 1977, 261.
62. **Hasegawa, M.,** Evolution of transfer RNA, in *Origin of Life,* Noda, H., Ed., Center for Academic Publications, Japan, 1978, 495.
63. **Hasegawa, M.,** Natural selection in the evolution of transfer RNA, in *Natural Selection,* Novak, V. J. A., Leonovich, V. V., and Pacltova, B., Eds., Czechoslovak Academy of Sciences, Prague, 1978, 663.
64. **Eigen, M. and Winkler-Oswatitsch, R.,** Transfer RNA, an early adaptor. Was it a primordial gene too? Manuscript, 1979.
64a. **Fitch, W. M.,** Distinguishing homologous from analogous proteins, *Syst. Zool.,* 19, 99, 1970.
65. **Schwartz, R. M., Gant, M. J., and Dayhoff, M. O.,** Transfer RNA, in *Atlas of Protein Sequence and Structure,* Vol. 5 (Suppl. 2), Dayhoff, M. O., Ed., National Biomedical Research Foundation, Washinton, D. C., 1976, 271.
66. **Raba, M., Limburg, K., Burghagen, M., Katze, J. R., Simsek, M., Heckman, J. E., RajBhandary, U. L., and Gross, H. J.,** Nucleotide sequence of three isoaccepting lysine tRNAs from a rabbit liver and SV-40 transformed mouse fibroblasts, *Eur. J. Biochem.,* 97, 305, 1979.
67. **Silverman, S., Gillam, I. C., Tener, G. M., and Söll, D.,** The nucleotide sequence of lysine $tRNA_2$ from *Drosophilia, Nucl. Ac. Res.,* 6, 435, 1979.
68. **Kimura, M. and Ohta, T.,** Eukaryotes-prokaryotes divergence estimates by 5S ribosomal RNA sequences, *Nature (London) New Biol.,* 243, 199, 1973.
69. **Nei, M.,** in *Molecular Population Genetics and Evolution,* Neuberger, A. and Fatum, E. L., Eds., North-Holland, Amsterdam, 1975, 72.
70. **Clarkson, S.,** Personal communication, 1979.
71. **Altwegg, M. and Kubli, E.,** The nucleotide sequence of phenylalanine $tRNA_2$ of *Drosophila melanogaster:* four isoacceptors with one basic sequence, *Nucl. Ac. Res.,* 6, 93, 1979.
72. **Barciszewski, J., Joachimiak, A., Rafalski, A., Barciszewska, M., Twardowski, T., and Wiewiorowski, Z.,** Conservation of the structures of plant tRNAs and aminoacyl-tRNA synthetases, *FEBS Lett.,* 102, 194, 1979.
73. **Yang, C. and Söll, D.,** Studies of transfer RNA tertiary structure by singlet-singlet energy transfer, *Proc. Natl. Acad. Sci. U.S.A.,* 71, 2838, 1974.
74. **Yaniv, B. G., Favre, A., and Barrell, B. G.,** Structure of transfer RNA, *Nature (London),* 223, 1331, 1969.
75. **Favre, A., Yaniv, A., and Michelson, A. M.,** The photochemistry of 4-thiouridine in *Escherichia coli* $tRNA_1^{Val}$, *Biochem. Biophys. Res. Commun.,* 37, 266, 1969.
76. **Krauskopf, M., Chen, C. M., and Ofengand, J.,** Interaction of fragmented and cross-linked *Escherichia coli* valine transfer ribonucleic acid with Tu factor-guanosine triphosphate complex, *J. Biol. Chem.,* 247, 842, 1972.
77. **Carré, D. S., Thomas, G., and Favre, A.,** Conformation and functioning of tRNAs: cross-linked tRNAs as substrate for tRNA nucleotidyl-transferase and aminoacyl synthetases, *Biochimie,* 56, 1089, 1974.
78. **Uhlenbeck, O. C.,** Complementary oligonucleotide binding to transfer RNA, *J. Mol. Biol.,* 65, 25, 1972.

79. **Reid, B. R. and Hurd, R. E.,** Application of high resolution nuclear magnetic resonance spectroscopy in the study of base pairing and the solution structure of transfer RNA, *Acct. Chem. Res.,* 10, 396, 1977.
80. **Clark, B. F. C. and Klug, A.,** The relationship of structure and function of tRNA with special reference to the three-dimensional structure of yeast phenylalanine tRNA, in *Proc. 10th FEBS Meet.,* Vol. 39, Chapeville, F. C. and Grundberg-Manago, M., Eds., American Elsevier, New York, 1975, 47.
81. **Brennan, T. and Sundaralingam, M.,** Structure of transfer RNA molecules containing the long variable loop, *Nucl. Ac. Res.,* 3, 3235, 1976.
82. **McCutchan, T., Silverman, S., Kohli, J., and Söll, D.,** Nucleotide sequence of phenylalanine transfer RNA from *Schizosaccharomyces pombe:* implications for transfer RNA recognition by yeast phenylalanyl tRNA synthetase, *Biochemistry,* 17, 1622, 1978.
83. **Dube, S. K.,** Recognition of tRNA by the ribosome. A possible role for 5S RNA, *FEBS Lett.,* 36, 39, 1973.
84. **Wrede, P., Woo, N. H., and Rich, A.,** Initiator tRNAs have a unique anticodon loop conformation, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 3289, 1979.
85. **Borer, P., Dengler, B., Tinoco, I., and Uhlenbeck, O.,** Stability of ribonucleic acid double stranded helices, *J. Mol. Biol.,* 86, 843, 1974.
86. **Schimmel, P. R. and Söll, D.,** Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs, *Annu. Rev. Biochem.,* 48, 601, 1979.
87. **Loftfield, R. B. and Vanderjagt, D.,** The frequency of errors in protein biosynthesis, *Biochem. J.,* 128, 1353, 1972.
88. **Barnett, W. E., Schwartzbach, S. D., and Hecker, L. I.,** The transfer RNAs of eukaryotic organelles, *Progr. Nucleic Acid Res. Mol. Biol.,* 18, 143, 1978.
89. **Winter, G. P. and Hartley, B. S.,** The amino acid sequence of tryptophanyl-tRNA synthetase from *Bacillus stearothermophilus, FEBS Lett.,* 80, 340, 1977.
90. **Koch, G. L. E., Boulanger, Y., and Hartley, B. S.,** Repeating sequences in aminoacyl-tRNA synthetases, *Nature (London),* 249, 316, 1974.
91. **Robbe-Saul, S., Fasiolo, F., and Boulanger, Y.,** Phenylalanyl-tRNA synthetase from baker's yeast. Repeated sequences in the two subunits, *FEBS Lett.,* 84, 57, 1977.
92. **Kula, M. R.,** Structural studies on isoleucyl-tRNA synthetase from *E. coli, FEBS Lett.,* 35, 299, 1973.
93. **Bruton, C. J., Jakes, P., and Koch, G. L. E.,** Repeated sequences in methionyl-tRNA synthetase from *E. coli, FEBS Lett.,* 45, 26, 1974.
94. **Kern, D., Potier, S. Boulanger, Y., and Lapointe, J.,** The monomeric glutamyl-tRNA synthetase of *Escherichia coli.* Purification and relation between its structural and catalytic properties, *J. Biol. Chem.,* 254, 518, 1979.
95. **Schimmel, P. R.,** Understanding the recognition of transfer RNA by aminoacyl-tRNA synthetases, *Adv. Enzymol. Relat. Areas Mol. Biol.,* 49, 187, 1979.
96. **Schimmel, P. R.,** Recent results on how aminoacyl transfer RNA synthetases recognize specific transfer RNAs, *Mol. Cell. Biochem.,* 25, 3, 1979.
97. **Igloi, G. L. and Cramer, F.,** Interaction of aminoacyl-tRNA synthetases and their substrates with a view to specificity, in *Transfer RNA,* Altman, S., Ed., MIT Press, Cambridge, 1978, 294.
98. **Yarus, M. and Mertes, M.,** The variety of intraspecific misacylations carried out by isoleucyl transfer ribonucleic acid synthetase of *Escherichia coli, J. Biol. Chem.,* 248, 6744, 1973.
99. **Ebel, J. P., Giegi, R., Bonnet, J., Kern, D., Befort, N., Bollack, C., Fasiolo, F., Gangloff, J., and Dirheimer, G.,** Factors determining the specificity of the tRNA-aminoacyl-tRNA synthetase recognition and particular importance of the maximal velocity, *Biochimie,* 55, 547, 1973.
100. **Crothers, D. M., Seno, T., and Söll, D.,** Is there a discriminator site in transfer RNA?, *Proc. Natl. Acad. Sci. U.S.A.,* 69, 3063, 1972.
101. **Wetzel, R.,** Aminoacyl-tRNA synthetase families and their significance to the origin of the genetic code, *Origins Life,* 9, 39, 1978.
102. **Sprinzl, M. and Cramer, F.,** Accepting site for aminoacylation of $tRNA^{Phe}$ from yeast, *Nature (London) New Biol.,* 245, 3, 1973.
103. **Hecht, S. M. and Chinault, C.,** Position of aminoacylation of individual *Escherichia coli* and yeast tRNAs, *Proc. Natl. Acad. Sci. U.S.A.,* 73, 405, 1976.
104. **Chinault, A. C., Hock Tan, K., Hassur, S. M., and Hecht, S. M.,** Initial position of aminoacylation of individual *Escherichia coli* yeast and calf liver transfer RNAs, *Biochemistry,* 16, 766, 1977.
105. **Lapointe, J.,** Personal communication.
106. **Martin, N., Rabinowitz, M., and Fukuhara, H.,** Isoaccepting mitochondrial glutamyl-tRNA species transcribed from different regions of the mitochondrial genome of *Saccharomyces cerevisiae, J. Mol. Biol.,* 101, 285, 1976.
107. **Nishimura, S.,** Modified nucleosides and isoaccepting tRNA, in *Transfer RNA,* Altman, S., Ed., MIT Press, Cambridge, 1978, chap. 7.

108. **Schäfer, K. P. and Söll, D.,** New aspects in tRNA biosynthesis, *Biochemie,* 56, 795, 1974.
109. **Katze, J. R. and Farkas, W. R.,** A factor in serum and aminiotic fluid is a substrate for the tRNA-modifying enzyme tRNA-guanine transferase, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 3271, 1979.
110. **Okada, N., Noguchi, S., Kasai, H., Shindo-Okada, N., Ohgi, T., Goto, T., and Nishimura, S.,** Novel mechanism of post-transcriptional modification of tRNA. Insertion of bases of Q precursors into tRNA by a specific tRNA transglycosylase reaction, *J. Biol. Chem.,* 254, 3067, 1979.
111. **McLennan, B. D.,** Enzymatic demodification of transfer RNA species containing n^6-(Δ^2-isopentenyl) adenosine, *Biochem. Biophys. Res. Commun.,* 65, 345, 1975.
112. **Feldman, M. Y.,** Minor components in transfer RNA: the location-function relationships, *Progr. Biophys. Mol. Biol.,* 32, 83, 1977.
112a. **Singhal, R. P. and Fallis, P. A. M.,** Structure, function and evolution of transfer RNAs, *Progr. Nucleic Acid Res. Mol. Biol.,* 23, 227, 1979.
113. **Cortese, R., Landsberg, R., Von der Haar, R. A., Umbarger, H. E., and Ames, B. N.,** Pleiotropy of His T mutants blocked in pseudouridine synthesis in tRNA: leucine and isoleucine-valine operons, *Proc. Natl. Acad. Sci. U.S.A.,* 71, 1857, 1974.
114. **Cedergren, R. J. and Cordeau, J. R.,** The distribution of modified nucleosides in transfer RNAs, *J. Theor. Biol.,* 39, 477, 1973.
115. **Nau, F.,** Species- and site-specificity of tRNA methylases: a logical approach, in *Biochemistry of Adenosylmethionine. Part 5. Polyamine Biosynthesis and Other Roles,* Columbia University Press, New York, 1977, 258.
116. **Carbon, J. and Fleck, E.,** Genetic alterations of structure and function in glycine tRNA of *E. coli:* mechanism of suppression of the tryptophan synthetase A78, *J. Mol. Biol.,* 85, 371, 1974.
117. **Turnbough, C. L., Neill, R. J., Landsberg, R., and Ames, B. N.,** Pseudouridylation of tRNAs and its role in regulation in *Salmonella typhimurium, J. Biol. Chem.,* 254, 5111, 1979.
118. **Hirsh, D. and Gold, L.,** Translation of the UGA triplet *in vitro* by tryptophan transfer RNA, *J. Mol. Biol.,* 58, 459, 1971.
119. **Hoburg, A., Aschoff, H. J., Kersten, H., Manderschied, F. R., and Gassen, H. G.,** On the function of the modified nucleosides m^7G, rT and ms^2i^6A in prokaryotic transfer RNA, *J. Bacteriol.,* 140, 408, 1979.
120. **Borek, E.,** Modification of nucleic acids in relation to differentiation, *Trends Biochem. Sci.,* 2, 3, 1977.
121. **Dingermann, T., Pistel, F., and Kersten, H.,** Functional role of ribosylthymine in transfer RNA, *Eur. J. Biochem.,* 104, 33, 1980.
122. **Fournier, M. J., Webb, E., and Kitchingham, G. R.,** General and specific effects of amino acid starvation on the formation of undermodified *Escherichia coli* phenylalanine tRNA, *Biochem. Biophys. Acta,* 454, 97, 1976.
123. **Chu-Der, O. M. Y. and Ortwerth, B. J.,** Lys-$tRNA_4$ and cell division. Changes in Lys-tRNA during the growth of mouse L cells, *Exp. Cell Res.,* 121, 291, 1979.
124. **Kuchino, Y. and Borek, E.,** Tumor specific Phe tRNA contains two supernumerary methylated bases, *Nature (London),* 271, 126, 1978.
125. **Razel, A. J. and Gray, E. D.,** Modification of phenylalanine tRNA during adaptation to anaerobic photosynthetic growth of *Rhodopseudomonas sphaeroides, Fed. Proc., Fed. Am. Soc. Exp. Biol.,* 36, 797, 1977.
126. **Roe, B. A., Chen, E. Y., Rizi, H. L., Stankiewicz, A. F., Weisz, C. C., Ma, D. P., Chen, C. Y., and Armstrong, P. W.,** The role of modified nucleotides in mammalian tRNA. The Q nucleoside is replaced by guanosine in a tumor $tRNA^{Asn}$, in *tRNA: Structure, Properties and Recognition,* Schimmel, P., Söll D., and Abelson, J., Eds., Cold Spring Harbor Laboratory, New York, 1980.
127. **Arnold, H. H., Raettig, R., and Keith, G.,** Isoaccepting phenylalanine tRNAs from *Bacillus subtilis* as a function of growth conditions. Differences in the content of modified nucleosides, *FEBS Lett.,* 73, 210, 1977.
128. **Sharma, O. K., Beezly, D. N., and Borek, E.,** Modulation of the synthesis *in vitro* of a harmone-induced protein by transfer RNA, *Nature (London),* 262, 62, 1976.
129. **Owenly, R. K., Stulberg, M. P., and Jacobson, B. K.,** Alteration of the Q Family of Transfer RNAs in adult *Drosophila* as a Function of Age, Nutrition and Genotype, *Mech. Aging Dev.,* 11, 91, 1979.
130. **Dingerman, T., Mach, M., and Kersten, H.,** Synthesis of Transfer Ribo-Nucleic Acids with Uridine or 2'-0-methylribothymidine at Position 54 in *Dictyosteluim discoideum, J. Gen. Microbiol.,* 115, 223, 1979.
131. **Borek, E., Gehrke, C. W., and Waalkes, T. P.,** Aberrant methylation of tRNA in tumor tissue, in *Transmethylation,* North-Holland/Elsevier, Amsterdam, 1979, 457.
132. **Kersten, H., Raettig, R., Weissenbach, J., and Dirheimer, G.,** Recognition of individual prokaryotic and eukaryotic transfer ribonucleic acids by *B. subtilis* adenine-1-methyltransferase specific for the dihydrouridine loop, *Nucl. Ac. Res.,* 5, 3033, 1978.
133. **Ninio, J.,** Approaches moléculaires de l'évolution, Masson, Paris, 1979, chap. III.

134. **Kubli, E.,** Transfer RNA modification in eukaryotes: an evolutionary interpretation, *Trends Biochem. Sci.,* 5, 190, 1980.
135. **Weissblum, B., Benzer, S., and Holley, R. W.,** A physical basis for the degeneracy of the amino acid code, *Proc. Natl. Acad. Sci. U.S.A.,* 48, 1449, 1962.
136. **Goodman, H. M., Abelson, J., Landy, A., Brennan, S., and Smith, J. D.,** Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA, *Nature (London),* 217, 1019, 1968.
137. **Celis, J. E., Hooper, M. L., and Smith, J. D.,** Aminio acid acceptor site of *E. coli* suppressor $tRNA^{Tyr}$ is a site of synthetase recognition, *Nature (London) New Biol.,* 244, 261, 1973.
138. **Nishimura, S.,** Minor components in transfer RNA: the components in transfer RNA: their characterization, location and function, *Progr. Nucleic Acid Res. Mol. Biol.,* 12, 49, 1972.
139. **Roberts, J. W. and Carbon, J.,** Molecular mechanism for missense suppression in *E. coli, Nature (London),* 250, 412, 1974.
140. **Grosjean, H. and Chantrenne, H.,** On codon-anticodon interactions, in *On Biological Recognition,* Chapeville, and Haenni, Eds., Springer-Verlag, 1979, in press.
141. **Gefter, M. L. and Russel, R. L.,** Role of modifications in tyrosine transfer RNA: a modified base affecting ribosome binding, *J. Mol. Biol.,* 39, 145, 1969.
142. **Miller, J. P., Hussain, Z., and Schweizer, M. P.,** The involvement of the anticodon adjacent modified nucleoside N-[9-(β-D-ribofuranosyl)purine-6-ylcarbamoyl]-threonine in the biological function of *E. coli* $tRNA^{Ile}$, *Nucl. Ac. Res.,* 3, 1185, 1976.
143. **Thiebe, R. and Zachau, H. G.,** A specific modification next to the anticodon of phenylalanine transfer ribonucleic acid, *Eur. J. Biochem.,* 5, 546, 1968.
144. **Ghosh, K. and Ghosh, H. P.,** Role of modified nucleosides in transfer ribonucleic acids. Effect of removal of the modified base adjacent to 3′ end of the anticodon on codon-anticodon interaction, *J. Biol. Chem.,* 247, 3369, 1972.
145. **Kimball, M. E. and Söll, D.,** The phenylalanine tRNA from *Mycoplasma sp.* (Kid): a tRNA lacking hypermodified nucleosides functional in protein synthesis, *Nucl. Ac. Res.,* 1, 1713, 1974.
146. **Litwack, M. D. and Peterkofsky, A.,** Transfer ribonucleic acid deficient in N^6-(Δ^2-isopentenyl) adenosine due to mevalonic acid limitation, *Biochemistry,* 10, 994, 1974.
147. **Laten, H., Gorman, J., and Bock, R. M.,** Isopentenyladenosine deficient tRNA from an antisuppressor mutant of *Saccharomyces cerivisiae, Nucl. Acid Res.,* 5, 4329, 1978.
148. **Jack, A., Ladner, J. E., and Klug, A.,** Crystallographic refinement of yeast phenylalanine transfer RNA at 2.5Å resolution, *J. Mol. Biol.,* 108, 619, 1976.
149. **Quigley, G. J., Wang, A. H. J., Seeman, N. C., Suddath, F. L., Rich, A., Sussman, J. L., and Kim, S. H.,** Hydrogen bonding in yeast phenylalanine transfer RNA, *Proc. Natl. Acad. Sci. U.S.A.,* 72, 4866, 1975.
150. **Kan, L. S., Ts'o, P. O. P., Sprinzl, M., von der Haar, F., and Cramer, F.,** ^1H nuclear magnetic resonance studies of transfer RNA: the methyl and methylene resonances of baker's yeast phenylalanine transfer RNA and its fragments, *Biochemistry,* 16, 3143, 1977.
151. **Langlois, R., Kim, S. H., and Cantor, C. R.,** A comparison of the fluorescence of the Y base of yeast $tRNA^{Phe}$ in solution and crystals, *Biochemistry,* 14, 2554, 1975.
152. **Fuller, W. and Hodgson, A.,** Conformation of the anticodon loop in tRNA, *Nature (London),* 215, 817, 1967.
153. **Eisinger, J., Fever, B., and Yamane, T.,** Codon-anticodon binding in $tRNA^{Phe}$, *Nature (London) New Biol.,* 231, 126, 1971.
154. **Grosjean, H., de Henau, S., and Crothers, D.,** On the physical basis for ambiguity in genetic coding, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 610, 1978.
155. **Grosjean, H., Söll, D. G., and Crothers, D. M.,** Studies of the complex between transfer RNAs with complementary anticodons. Origins of enhanced affinity between complementary triplets, *J. Mol. Biol.,* 103, 499, 1976.
156. **Weissenbach, J. and Grosjean, H.,** Role of hypermodified bases in anticodon of tRNA: effect of N-(9-β-D ribofuranosyl purin-6-yl) carbamoyl-threonine (t^6A) on efficiency of codon-anticodon and anticodon-anticodon interactions, Manuscript, 1980.
157. **Steege, D. A.,** 5′-Terminal nucleotide sequence of *E. coli* lactose repressor RNA: features of translational initiation and reinitiation sites, *Proc. Natl. Acad. Sci. U.S.A.,* 74, 4163, 1977.
158. **Yarus, M.,** The accuracy of translation, *Progr. Nucleic Acid Res. Mol. Biol.,* 21, 195, 1979.
159. **Grosjean, H. and Buckingham, R.,** Accuracy of protein synthesis: a reexamination of specificity in codon-anticodon interaction, *Progr. Nucleic Acid Res. Mol. Biol.,* in press.
160. **Kurland, C. G., Rigler, R., Ehrenberg, M., and Blomberg, C.,** Allosteric mechanism for codon-dependent tRNA selection on ribosomes, *Proc. Natl. Acad. Sci. U.S.A.,* 72, 4248, 1975.
161. **Hopfield, J. J.,** Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence and kinetic proofreading, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 4334, 1978.
162. **Ninio, J.,** Evaluation of some molecular parameters of translation *in vivo, J. Mol. Biol.,* 84, 297, 1974.

163. **Akaboschi, E., Inoye, M., and Tsugita, A.,** Effect of neighboring nucleotide sequences on suppression efficiency in amber mutants of T_4 phage lysozyme, *Mol. Gen. Genet.,* 149, 1, 1976.
164. **Colby, D. S., Schedl, P., and Guthrie, C.,** A functional requirement for modification of the wobble nucleotide in the anticodon of a T_4 suppressor tRNA, *Cell,* 9, 449, 1976.
165. **Feinstein, S. I. and Altman, S.,** Context effects on nonsense codon suppression in *E. coli, Genetics,* 88, 201, 1978.
166. **Sekiya, T., Takeishi, K., and Ukita, T.,** Specificity of yeast glutamic acid transfer RNA for codon recognition, *Biochim. Biophys. Acta,* 182, 411, 1969.
167. **Takemoto, T., Takeishi, K., Nishimura, S., and Ukita, T.,** Transfer of valine into rabbit haemoglobin from various isoaccepting species of valyl-tRNA differing in codon recognition, *Eur. J. Biochem.,* 38, 489, 1973.
168. **Hillen, W., Egert, E., Lindner, H. J., and Gassen, H. G.,** Crystal and molecular structure of 2-thio-5-carbonymethyluridine and its methyl ester: helix terminator nucleosides in the first position of some anticodons, *Biochemistry,* 17, 5314, 1978.
169. **Mitra, S. K., Lustig, F., Åkesson, B., Axberg, T., Elias, P., and Lagerkvist, U.,** Relative efficiency of anticodons in reading the valine codons during protein synthesis *in vitro, J. Biol. Chem.,* 254, 6397, 1979.
170. **Ninio, J.,** Codon-anticodon recognition: the missing triplet hypothesis, *J. Mol. Biol.,* 56, 63, 1971.
171. **Woese, C.,** Molecular mechanics of translation: a reciprocating ratchet mechanism, *Nature (London),* 226, 817, 1970.
172. **Labuda, D. and Pörske, D.,** Multistep mechanism of codon recognition by tRNA, *Biochemistry,* 19, 3799, 1980.
173. **Möller, A., Wild, U., Riesner, D., and Gassen, H. G.,** Evidence from ultraviolet absorbance measurement for a codon-induced conformational change in lysine tRNA from *Escherichia coli, Proc. Natl. Acad. Sci. U.S.A.,* 76, 3266, 1979.
174. **Wagner, R. and Garrett, R. L.,** Chemical evidence for a codon-induced allosteric change in $tRNA^{Lys}$ involving the 7-methylguanosine residue 46, *Eur. J. Biochem.,* 97, 615, 1979.
175. **Geerdes, H. A. M.,** NMR Study on tRNA Structure and Codon-Anticodon Interaction, Thesis, University of Nijmegen, the Netherlands, 1979.
176. **Shine, J. and Dalgarno, L.,** The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplet and ribosome binding sites, *Proc. Natl. Acad. Sci. U.S.A.,* 71, 1342, 1974.
177. **Berquist, P. L., Burns, D. J. W., and Plinston, C. A.,** Participation of redundant transfer ribonucleic acids from yeast in protein synthesis, *Biochemistry,* 7, 1751, 1968.
178. **Mitra, S. R., Lustig, F., and Åkesson, B., and Lagerkvist, U.,** Codon-anticodon recognition in the valine codon family, *J. Biol. Chem.,* 252, 471, 1977.
179. **Buckingham, R. H. and Kurland, C. G.,** Codon specificity of UGA suppressor $tRNA^{Trp}$ from *E. coli, Proc. Natl. Acad. Sci. U.S.A.,* 74, 5496, 1977.
180. **Goldman, E., Holmes, W. M., and Hatfield, G. W.,** Specificity of codon recognition by *E. coli* tRNA-leu isoaccepting species determined by protein synthesis *in vitro* directed by phage RNA, *J. Mol. Biol.,* 129, 567, 1979.
181. **Parker, J., Pollard, J. W., Friesen, J. D., and Stanners, C. P.,** Stuttering: high-level mistranslation in animal and bacterial cells, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 1091, 1978.
182. **O'Farrell, P. H.,** The suppression of defective translation by ppGpp and its role in the stringent response, *Cell,* 14, 545, 1978.
183. **Lagerkvist, U.,** Two out of three: an alternative method for codon reading, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 1759, 1978.
184. **Jukes, T. H.,** Possibilities for the evolution of the genetic code from a preceding form, *Nature (London),* 246, 22, 1973.
185. **Edelman, N. P. and Gallant, J.,** Mistranslation in *E. Coli, Cell,* 10, 131, 1977.
186. **Grosjean, H., Sankoff, D., MinJou, W., Fiers, W., and Cedergren, R. J.,** Bacteriophage MS2 RNA: a correlation between the stability of the codon-anticodon interaction and the choice of code words, *J. Mol. Evol.,* 12, 113, 1978.
187. **Grantham, R.,** Viral, prokaryote and eukaryote genes contrasted by mRNA sequence indexes, *FEBS Lett.,* 95, 1, 1978.
188. **Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pané, A.,** Codon catalog usage and the genome hypothesis, *Nucl. Acid Res.,* 8, r49, 1980.
189. **Grosjean, H.,** Codon usage in several organisms, in *tRNA: Biological Aspects,* Söll, D., Abelson, J., and Schimmel, P., Eds., Cold Spring Harbor Laboratory, New York, 1980.
190. **Kafatos, F. C., Efstratiadis, A., Forget, B. G., and Weissman, S. M.,** Molecular evolution of human and rabbit β-globulin mRNAs, *Proc. Natl. Acad. Sci. U.S.A.,* 74, 5618, 1977.
191. **Fiers, W. and Grosjean, H.,** On codon usage, *Nature (London),* 277, 328, 1979.

192. **Hensgens, L. A. M., Grivell, L. A., Borst, P., and Bos, J. L.,** Nucleotide sequence of the mitochondrial structural gene for subunit 9 of yeast ATPase complex, *Proc. Natl. Acad. Sci. U.S.A.*, 76, 1663, 1979.
193. **Macino, G. and Tzagoloff, A.,** Assembly of the mitochondrial membrane system. The DNA sequence of a mitochondrial ATPase gene in *Saccharomyces cerevisiae*, *J. Biol. Chem.*, 254, 4617, 1979.
194. **McClain, W. H., Guthrie, C., and Barrel, B. G.,** Eight transfer RNAs induced by infection of *Escherichia coli* with bacteriophage T_4, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 3703, 1972.
194a. **McClain, W. H.,** Personal communication.
195. **Yudelevitch, A.,** Specific cleavage of an *E. coli* leucine transfer RNA following bacteriophage T^4 infection, *J. Mol. Biol.*, 60, 21, 1971.
196. **Pinkerton, T. C., Paddock, G., and Abelson, J.,** Bacteriophage T_4 $tRNA^{Leu}$, *Nature (London) New Biol.*, 240, 88, 1972.
196a. **Elton, R. A., Russell, G. J., and Subak-Sharpe, J. H.,** Doublet frequencies and codon weighting in the DNA of *E. coli* and its phages, *J. Mol. Evol.*, 8, 117, 1976.
196b. **Scherberg, N. H. and Weiss, S. B.,** T_4 transfer RNAs: codon recognition and translational properties, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 1114, 1972.
197. **Abelson, J., Fukada, K., Johnson, P., Lamfrom, H., Nierlich, D. P., Otsuka, A., Paddock, G. V., Pinkerton, T. C., Sarabhai, A., Stahl, S., Wilson, J. H., and Yesian, H.,** Bacteriophage T_4 tRNAs: structure, genetics and biosynthesis, in *Brookhaven Symp. Biol. No. 26, Processing of RNA*, Dunn, J. J., Ed., Brookhaven National Laboratory, 1975, 77.
198. **Garel, J. P.,** Quantitative adaptation of isoacceptor tRNAs to mRNA codons of alanine, glycine and serine, *Nature (London)*, 260, 805, 1976.
198a. **Chavancy, G., Chevallier, A., Fournier, A., and Garel, J. P.,** Adaptation of iso-tRNA concentration to mRNA codon frequency, *Biochimie*, 61, 7, 1979.
199. **Abelson, J.,** RNA processing and the intervening sequence problem, *Annu. Rev. Biochem.*, 48, 1035, 1979.
200. **Miller, J. R. and Brownlee, G. G.,** Is there a correction mechanism in the 5S multigene system, *Nature (London)*, 275, 556, 1978.
201. **McClain, W. H., Seidman, J. G., and Schmidt, F.,** Evolution of the biosynthesis of the 3'-terminal CCA residues in T-even bacteriophage transfer RNAs, *J. Mol. Biol.*, 119, 519, 1978.
202. **Bos, J. L., Osinga, K. A., van der Horst, G., and Borst, P.,** Nucleotide sequence of the mitochondrial structural genes for cysteine-tRNA and histidine-tRNA of yeast, *Nucl. Ac. Res.*, 6, 3255, 1979.
202a. **Goodman, H. M., Olson, M. V., and Hall, B. D.,** Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor *SUP4-0*, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 5453, 1977.
203. **Altman, S. and Smith, J. D.,** Tyrosine tRNA precursor molecule polynucleotide sequence, *Nature (London) New Biol.*, 233, 35, 1971.
204. **Chang, S. and Carbon, J.,** The nucleotide sequence of a precursor to the glycine- and threonine-specific transfer ribonucleic acids of *E. coli*, *J. Biol. Chem.*, 250, 5542, 1975.
205. **Reanney, D. C.,** On the origin of prokaryotes, *J. Theor. Biol.*, 48, 243, 1974.
206. **Dillon, L. S.,** *The Genetic Code Mechanism and the Origin of Life*, Plenum Press, New York, 1978.
207. **Zablen, L. B., Hissil, M. S., Woese, C. R., and Buetow, D. E.,** Phylogenetic origin of the chloroplast and prokaryotic nature of its ribosomal RNA, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 2418, 1975.
208. **Bonen, L. and Doolittle, W. F.,** Partial sequences of 16S RNA and the phylogeny of blue-green algae and chloroplasts, *Nature (London)*, 261, 669, 1976.
209. **Raff, R. A. and Mahler, H. R.,** Origin of mitochondria, *Science*, 180, 516, 1973.
210. **Bonen, L., Cunningham, R. S., Gray, M. W., and Doolittle, W. F.,** Wheat embryo mitochondrial 18S ribosomal RNA: evidence for its prokaryotic nature, *Nucl. Ac. Res.*, 4, 663, 1977.
211. **Woese, C. R.,** Endosymbionts and mitochondrial origins, *J. Mol. Evol.*, 10, 93, 1977.
212. **Tzagoloff, A., Macino, G., and Sebald, W.,** Mitochondrial genes and translation products, *Annu. Rev. Biochem.*, 48, 419, 1979.
213. **Buetow, D. E. and Wood, W. M.,** The Mitochondrial Translation System, *Subcell. Biochem.*, 5, 1, 1978.
214. **Bernardi, G.,** Organization and evolution of the mitochondrial genome of yeast, *J. Mol. Evol.*, 9, 25, 1976.
215. **Jakovcic, S., Hendler, F., Halbreich, A., and Rabinowitz, M.,** Transcription of yeast mitochondrial deoxyribonucleic acid, *Biochemistry*, 18, 3200, 1979.
216. **Bernardi, G.,** The "petite" mutation in yeast, *Trends Biochem. Sci.*, 4, 197, 1979.
217. **Groot, S. P., Flavell, R. A., and Sanders, J. P. M.,** Sequence homology of nuclear and mitochondrial DNAs of different yeasts, *Biochim. Biophys. Acta*, 378, 186, 1975.
218. **de Vries, H., de Jonge, J. C., Bakker, H., Meurs, H., and Kroon, A.,** The anatomy of the tRNA-rRNA region of the *Neurospora crassa* mitochondrial DNA, *Nucl. Ac. Res.*, 6, 1791, 1979.

219. **Doolittle, W. F.,** Genes in pieces: were they ever together, *Nature (London),* 272, 581, 1978.
220. **Heckman, J. E., Hecker, L. I., Schwartzbach, S. D., Barnett, W. E., Baumstark, B., and RajBhandary, U. L.,** Structure and function of initiator methionine tRNA from the mitochondria of *Neurospora crassa, Cell,* 13, 83, 1978.
221. **Heckman, J. E., Alzner-Deweerd, B., and RajBhandary, U. L.,** Interesting and unusual features in the sequence of *Neurospora crassa* mitochondrial tyrosine transfer RNA, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 717, 1979.
222. **Li, M. and Tzagoloff, A.,** Assembly of the mitochondrial membrane system: sequences of yeast mitochondrial valine and an unusual threonine tRNA gene, *Cell,* 18, 47, 1979.
223. **Miller, D. L., Martin, N. C., Pham, H. D., and Donelson, J. E.,** Sequence analysis of two yeast mitochondrial DNA fragments containing the genes for $tRNA^{Ser}_{UCR}$ and $tRNA^{Phe}_{UUY}$, *J. Biol. Chem.* 254, 11735, 1979.
224. **Martin, N. C., Miller, D. L., Donelson, J. E., Sigurdson, C., Hartley, J. L., Moynihan, P. S., and Pham, H. D.,** Identification and sequencing of yeast mitochondrial tRNA genes in mitochondrial DNA-pBR322 recombinants, in *Extrachromosomal DNA,* ICN-UCLA Symp. Mol. Cell. Biol. 15, Cummings, D., Borst, P., Davis, J., Weissman, S., and Fox, C. F., Eds., Academic Press, New York, in press.
225. **Miller, D. L., Sigurdson, C., Martin, N. C., and Donelson, J. E.,** Nucleotide sequence of the mitochondrial genes coding for $tRNA^{Gly}_{GGR}$ and $tRNA^{Val}_{GUR}$, *Nucl. Ac. Res.,* 8, 1435, 1980.
226. **Barrell, B. G., Bankier, A. T., and Drouin, J.,** A different genetic code in human mitochondria, *Nature (London),* 282, 189, 1979.
227. **Heckman, J. E., Sarnoff, J., Alzer-DeWeerd, B., Yin, S., and RajBhandary, U. L.,** Novel features in the genetic code and codon reading patterns in *N. crassa* based on sequences of six mitochondrial tRNAs, *Proc. Natl. Acad. Sci. U.S.A.,* 77, 3159, 1980.
228. **Walker, R. T. and RajBhandary, U. L.,** The nucleotide sequence of formylmethionine tRNA from *Mycoplasma mycoides sp. capri, Nucl. Ac. Res.,* 5, 57, 1978.
229. **Macino, G., Coruzzi, G., Nobrega, F. G., Li, M., and Tzagoloff, A.,** Use of the UGA terminator as a tryptophan codon in yeast mitochondria, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 3784, 1979.
230. **Martin, R. P., Schneller, J. M., Sibbler, A. P., Stahl, A. J. C., and Dirheimer, G.,** Nuclear-DNA-coded $tRNA^{Lys}$ (anticodon CUU) is the only cytoplasmic tRNA imported into yeast mitochondria, *Biochemistry,* 18, 4600, 1979.
231. **Chiu, N., Chiu, A., and Suyama, Y.,** Native and imported transfer RNA in mitochondria, *J. Mol. Biol.,* 99, 37, 1975.
232. **Ninio, J.,** Considerations on the problem of the joint evolution of two different translation apparatuses within the same cells, in *Molecular Biology of Nucleocytoplasmic Relationships,* Elsevier, Amsterdam, 1975, 31.
233. **Woese, C. R. and Bleyman, M. A.,** Genetic code limit organisms — do they exist?, *J. Mol. Evol.,* 1, 223, 1972.
234. **Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R., and Rutter, W.,** Structure of yeast phenylalanine-tRNA genes: an intervening DNA segment within the region coding for the tRNA, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 190, 1978.
235. **Yot, P., Pinck, M., Haenni, A. L., Duranton, H. M., and Chapeville, F.,** Valine-specific tRNA-like structure in turnip yellow mosaic virus RNA, *Proc. Natl. Acad. Sci. U.S.A.,* 67, 1345, 1970.
236. **Haenni, A. L., Joshi, S., Hubert, E., Huez, G., and Marbaix, G.,** *In vivo* aminoacylation and processing of turnip-yellow mosaic virus RNA injected into *X. laevis* Oocytes, in *tRNA, Biological Aspects,* Söll, D., Abelson, J., and Schimmel, P., Eds., Cold Spring Harbor Laboratory, New York, 1980.
237. **Crick, F. H. C., Brenner, S., Klug, A., and Pieczenik, G.,** A speculation on the origin of protein synthesis, *Origins Life,* 7, 389, 1976.
238. **Ninio, J.,** Approches *Moléculaires de l'Evolution,* Masson, Paris, 1979, 83.
239. **Ehran, S., Greller, L. D., and Rasco, B.,** Evolution of the transfer RNA molecule, *Z. Naturforsch.,* 32C, 413, 1977.
240. **Cedergren, R. J.,** unpublished data.
241. **Barrell, B. G., Anderson, S., Bankier, A. T., de Bruijn, H. L., Chen, E., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., and Young, I. G.,** Different pattern of codon recognition by mammalian mitochondrial tRNAs, *Proc. Natl. Acad. Sci. U.S.A.,* 77, 3164, 1980.
242. **Bonitz, S. G., Berlani, R., Coruzzi, G., Li, M., Macino, G., Norbrega, F. G., Norbrega, M. P., Thalenfeld, B. E., and Tzagoloff, A.,** Codon recognition rules in yeast mitochondria, *Proc. Natl. Acad. Sci.,* 77, 3167, 1980.

243. **Grosjean, H., de Henau, S., Sibler, A. P., Martin, R., Dirheimer, G., Keith, G., and Kohli, J.,** Mitochondria tryptophanyl-tRNA^{Trp} from yeast functions as a supersuppressor of UGA opal codons when macroinjected into cytoplasms of xenopus laevis oocyte, *Arch. Intern. Physiol. Biochem.*, in press.

244. **Arcari, P., and Brownlee, G. G.,** The nucleotide sequence of a small (3S) serly-tRNA (anticodon GCU) from beef heart mitochondria, *Nucl. Acid. Res.*, 8, 5213, 1980.

245. **De Bruijn, M. H. L., Schreier, P. H., Eperon, I. C., Barrell, B. G., Chen, E. Y., Armstrong, P. W., Wong, J. F. H., and Roe, B. A.,** A mammalian mitochondrial serine tRNA lacking the "dihydrourine" loop and stem, *Nucl. Acid Res.*, 8, 5213, 1980.