J Mol Evol (1995) 40:509–518

JOURNAL OF MOLECULAR EVOLUTION

Evolution of tRNA Recognition Systems and tRNA Gene Sequences

Margaret E. Saks, Jeffrey R. Sampson

Division of Biology 147-75, California Institute of Technology, Pasadena CA 91125, USA

Received: 27 September 1994

Abstract. The aminoacylation of tRNAs by the aminoacyl-tRNA synthetases recapitulates the genetic code by dictating the association between amino acids and tRNA anticodons. The sequences of tRNAs were analyzed to investigate the nature of primordial recognition systems and to make inferences about the evolution of tRNA gene sequences and the evolution of the genetic code. Evidence is presented that primordial synthetases recognized acceptor stem nucleotides prior to the establishment of the three major phylogenetic lineages. However, acceptor stem sequences probably did not achieve a level of sequence diversity sufficient to faithfully specify the anticodon assignments of all 20 amino acids. This putative bottleneck in the evolution of the genetic code may have been alleviated by the advent of anticodon recognition. A phylogenetic analysis of tRNA gene sequences from the deep Archaea revealed groups that are united by sequence motifs which are located within a region of the tRNA that is involved in determining its tertiary structure. An association between the third anticodon nucleotide (N36) and these sequence motifs suggests that a tRNA-like structure existed close to the time that amino acid–anticodon assignments were being established. The sequence analysis also revealed that tRNA genes may evolve by anticodon mutations that recruit tRNAs from one isoaccepting group to another. Thus tRNA gene evolution may not always be monophyletic with respect to each isoaccepting group.

Key words: Genetic code — tRNA recognition — Gene recruitment — tRNA gene evolution

Introduction

Transfer RNA (tRNA) constitutes a family of molecules with similar sizes and tertiary structures (Fig. 1) that play a pivotal role in the execution of the genetic code and in the maintenance of translational fidelity. The family can be subdivided into 20 (isoaccepting) groups that are defined on the basis of the amino acid that the cognate aminoacyl-tRNA synthetase attaches to the tRNA 3′-terminus and by the sequence of the tRNA anticodon that engages in complementary base pairing with the codons of messenger RNA (mRNA). Although tRNAs within each isoaccepting group are aminoacylated with the same amino acid, members of each isoaccepting group often have different gene sequences. Within each tRNA sequence there exist elements that are unambiguously recognized by the cognate synthetase and that therefore allow synthetases to correctly aminoacylate their cognate tRNAs and to avoid misacylating tRNAs from the 19 non-cognate groups. In the tRNAs of extant organisms, these recognition elements are most commonly located in the tRNA anticodon, the acceptor stem and the associated single-stranded "discriminator" base at position 73 (Pallanck and Schulman 1992; Giegé et al. 1993; McClain 1993; Saks et al. 1994). In addition, synthetases sometimes recognize nucleotides in the tRNA variable pocket and aspects of tRNA structure. Although much attention has focused on elucidating the elements that account for the correct recognition of tRNAs in extant organisms, considerably less is known about the nature of primordial recognition systems and the steps that ac-

Based on a presentation made at a workshop—"Aminoacyl-tRNA Synthetases and the Evolution of the Genetic Code"—held at Berkeley, CA, July 17–20, 1994

Correspondence to: M.E. Saks

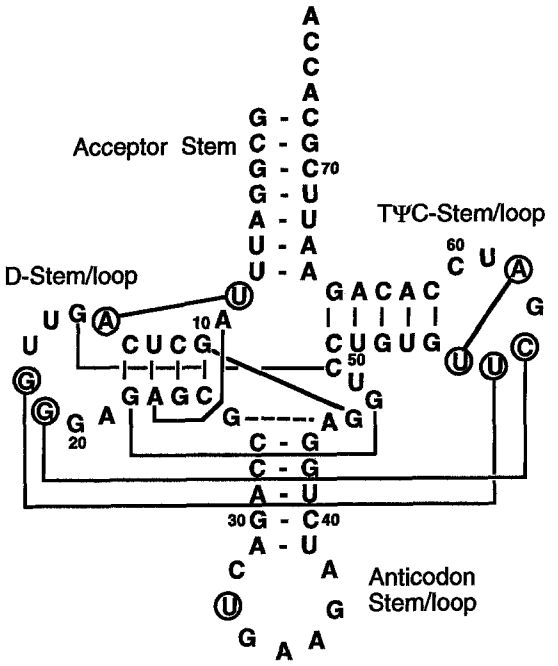

Fig. 1. Secondary structure of yeast $tRNA^{Phe}$ with tertiary interactions and base triples indicated. The tRNA sequence is presented without modified nucleotides in the standard cloverleaf format using the conventional numbering system (Gauss et al. 1979). *Circled* nucleotides denote those that are conserved among all nonmitochondrial tRNAs. *Solid lines* between noncontiguous nucleotides denote base pairs, tertiary base pairs, and base triple interactions. The *dashed line* indicates a nonstandard base pair that contributes to tertiary structure.

count for the evolution of extant recognition systems from these ancestors.

A number of different types of tRNA recognition systems were probably tried during and subsequent to the evolution of the amino acid–anticodon assignments that comprise the genetic code. Their success surely depended on whether they could ensure a correspondence between a particular tRNA anticodon and the amino acid attached to the tRNA 3′-terminus. On the other hand, the evolution of recognition systems was probably constrained by the variation that existed in tRNA gene sequences with respect to the types of tRNA structures that could be made and the diversity of information that could be presented for recognition by the interacting synthetases. An additional constraint was also probably imposed by the level of sophistication of the synthetases with regard to their ability to specifically recognize ribonucleotides as well as the structures of RNAs.

The problem of whether the L-shaped structure of extant tRNAs, having an anticodon stem/loop positioned at a 90° angle with respect to the 3′-terminus, evolved before or after the advent of templated protein synthesis is difficult. By definition, the earliest RNAs to participate in templated protein synthesis must have had a 3′-terminus to which an amino acid was attached and a single-stranded group of nucleotides capable of coupling with an mRNA. Perhaps the primordial tRNA-like structure evolved from small aminoacylatable RNA hairpins that first acquired an anticodon loop and later acquired a tRNA-like structure as they coevolved with the ribosome (Möller and Janssen 1990, 1992; Maizels and Weiner 1994). On the other hand, Weiner and Maizels (1987) have argued, based on considerations of the roles that tRNAs play in biological processes other than translation, that a tRNA-like structure evolved as a tag for RNA replication and was later recruited for templated protein synthesis. For the purpose of our exploration of the nature of ancient recognition systems, we have assumed that the primordial tRNA-like structures that participated in templated protein synthesis had at least a 3′-terminus for amino acid attachment and some sort of single-stranded loop capable of base pairing with mRNA codons. To explore the problem of when the tRNA-like structure evolved relative to the assignments of amino acids to anticodons, we analyzed the tRNA sequences of the Archaea.

Because the determinants for the correct aminoacylation of tRNAs are embedded within each tRNA gene sequence, these sequences should carry some record of the steps which eventually culminated in today's recognition systems. We therefore analyzed tRNA sequences from organisms representing the major phylogenetic lineages to look for enduring traces of tRNA recognition systems. In addition, we focused our attention on the tRNA sequences of ancient organisms since these are likely to give the closest possible approximation of the nature of primordial tRNAs and recognition systems. The overall goal was to explore the nature of primordial recognition systems, the factors dictating their evolution, and the ways in which different types of recognition systems may, in turn, influence tRNA gene evolution.

Results and Discussion

The Location of Recognition Elements in Primordial tRNAs

The ability of some present-day synthetases to specifically aminoacylate RNAs representing only the coaxial acceptor-TΨC stem implicates the acceptor stem as an ancient location for recognition elements (Schimmel et al. 1993). In addition, the lack of sophistication of the original synthetases and tRNAs may have constrained recognition elements to be in close proximity to the synthetase active site when the amino acid was transferred to the tRNA 3′-terminus. Based on this type of reasoning, the acceptor stem is a more attractive candidate for the location of recognition elements in primordial tRNAs than is the anticodon.

To explore the possible involvement of acceptor stem nucleotides in determining the amino acid identity of primordial tRNAs we analyzed all of the available nonmitochondrial tRNA sequences (Steinberg et al. 1993). We first determined, for each of the 20 isoaccepting groups, whether any acceptor stem nucleotides were con-

Table 1. Acceptor stem nucleotides conserved among tRNAs (non-mitochondrial) from the Bacteria, Archaea, and Eucarya[a]

	Acceptor stem position							
Isoacceptor	73	1-72	2-71	3-70	4-69	5-68	6-67	7-66
Ala (61)	A	G-C	G-C	G · U	—	—	—	
Arg (65)	—	—	—	—	—	—	—	—
Asn (35)	G	—	—	C-G	—	—	—	—
Asp (35)	G	—	—	—	—	—	—	—
Cys (20)	U	G-C	—	—	—	—	—	—
Glu (46)	—	—	—	C-G[b]	—	—	—	—
Gln (46)	—	—	G-C	—	—	—	—	—
Gly (63)	—	G-C	C-G	—	—	—	—	—
His (31)	BP	G-C	—	—	—	—	—	—
Ile (54)	A	G-C	G-C[b]	—	—	—	—	—
Leu (91)	A	G-C	—	—	—	—	—	—
Lys (50)	—	G-C	—	—	—	—	—	—
Met (34)	A	—	—	—	—	—	—	—
Phe (44)	A	G-C	—	—	—	—	—	—
Pro (48)	—	—	G-C	—	—	—	—	—
Ser (85)	—	G-C	—	—	—	—	—	—
Thr (51)	—	G-C	C-G[b]	—	—	—	—	—
Trp (27)	—	—	—	—	—	—	—	—
Tyr (41)	A	—	—	—	—	—	—	—
Val (60)	A	—	—	—	—	—	—	—

[a] Nucleotides that occur in 90% or more of the sequences for each isoaccepting group are presented. The number of tRNA sequences analyzed for each isoaccepting group is indicated in parentheses. The sequences were taken from the tRNA data base of Sprinzl and co-workers (Steinberg et al. 1993) and updated with sequences from GenBank. The GenBank accession numbers are as follows: X59857, X52382, S42235, S42231, S42232, M97643, M97642, M97641, M97644, X07692, X68397, L07299. The sequences for the genus *Pyrobaculum* were kindly provided by Ron Swanson. The *Sulfolobus solfataricus* $tRNA^{Ile}$ (CAU) is incorrectly identified as an initiator $tRNA^{Met}$(X0860) in the tRNA data base. BP = base pair.

[b] Nucleotides are conserved among 87–89% of the sequences.

served across the three major groups (Bacteria, Archaea, Eucarya; Woese 1987; Iwabe et al. 1989; Woese et al. 1990). If acceptor stem nucleotides made early and important contributions to recognition, their nucleotide identities and locations should have been conserved by strong stabilizing selection even as more complex recognition systems and organisms evolved. We then used the results of this analysis and an analysis of tRNAs from the Archaea to explore whether there was sufficient diversity in acceptor stem sequences to support a genetic code encompassing all 20 amino acids.

In nearly all of the 20 isoaccepting groups, at least one acceptor stem nucleotide has been conserved across the three organismal groups (Table 1). Thus there appears to have been strong stabilizing selection for the maintenance of these elements in each isoaccepting group. The conserved nucleotides are found between position 73 and the 3-70 base pair and thus are restricted to the upper portion of the acceptor stem helix. If the search is confined to closely related organisms, additional conserved nucleotides are revealed for each isoaccepting group. However, they are still generally located between position 73 and the 3-70 base pair. The propensity for base pairs near the top of the acceptor stem to be conserved across the three major organismal groups supports the idea that recognition sites were located in the acceptor stems of primordial tRNAs even before the major organismal groups diverged.

The analysis also revealed that tRNAs from different isoaccepting groups often have the same nucleotide or base pair conserved at a particular position within the acceptor stem. For example, an A73 is found in seven different isoaccepting groups. Moreover, G-C is the only base pair that is conserved at position 1-72 and it is found in ten isoaccepting groups. Thus although it is likely that acceptor stem nucleotides contributed to the recognition of primordial tRNAs, it is unlikely that the nucleotide or base pair at a single position within the acceptor stem was the sole determinant of tRNA amino acid specificity. Alanine and histidine tRNAs present the only likely exception. These two isoaccepting groups can be defined by a distinctive acceptor stem base pair that has been conserved throughout the evolution of the major organismal groups; and in both cases the base pair has been shown, experimentally, to make a very large contribution to aminoacylation by the cognate synthetase (Hou and Schimmel 1988; McClain and Foss 1988; Himeno et al. 1989; Francklyn and Schimmel 1990).

To answer the question of whether there was enough diversity in acceptor stem sequences to support a genetic code encompassing all 20 amino acids, we analyzed the sequences of tRNAs from the Archaea. The Crenarchaeota and the Euryarchaeota (*Methanopyrus, Thermococcus,* and *Methanococcus*) connect deep within the phylogeny revealed by Woese's analysis of 16S ribosomal RNA and have short branches (Woese et al. 1990; Olsen et al. 1994). Thus the tRNA sequences of these organisms provide a basis for pondering the characteristics of ancient tRNAs. By lumping organisms from the above archaeal groups, we obtained a data set representing all isoaccepting groups except tryptophan and cysteine. For the acceptor stem analysis, the sequences of $tRNA^{Trp}$ and $tRNA^{Cys}$ from *Halobacterium volcanii* were included since preliminary analyses of tRNAs from the Archaea indicated that there is little among-species variation with respect to the sequence at the top of the acceptor stem for each isoaccepting group. We concentrated on the region between position 73 and the 3-70 base pair for the analysis of the deep Archaea because our analysis of conserved acceptor stem nucleotides (Table 1) indicated that nucleotides in this region were most likely to have been involved in the recognition of ancient tRNAs.

There were only 11 different acceptor stem sequences among the 20 isoaccepting groups in the deep Archaea (Fig. 2). The sequence diversity is increased if all seven acceptor stem positions are included. However, because much of the added variation occurs within rather than among isoaccepting groups, it is not likely to help in discriminating one group from another (except when

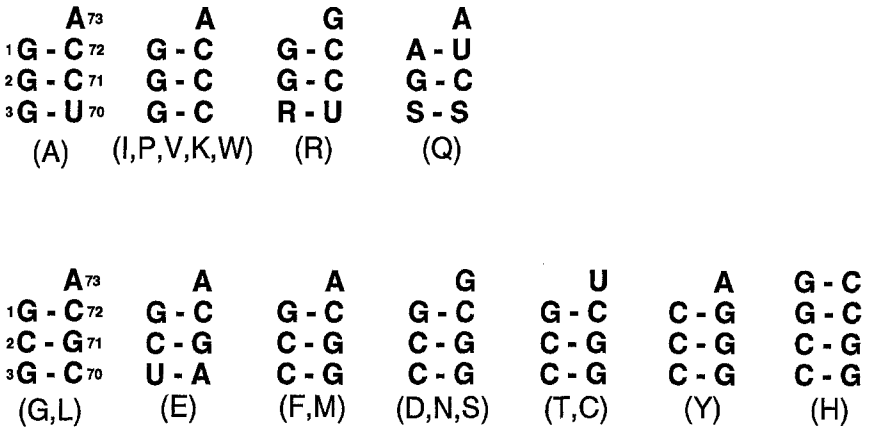

Fig. 2. The sequences at position 73 and the top three acceptor stem base pairs in deep Archaea tRNAs. The isoaccepting groups corresponding to each acceptor stem are given in *parentheses* using the single-letter amino acid abbreviations. Lysine tRNAs have either A or G at position 73; leucine and valine tRNAs have either G-C or A-U at position 3–70. See text and Fig. 3 legend for the sequences included in the analysis. Nucleotide abbreviations are R = A or G; S = G or C.

isoacceptors have different base pairs with a shared nucleotide chemistry that can be detected by the cognate synthetase; unpublished result). Thus, the diversity of acceptor stem sequences of these ancient organisms does not seem great enough to specify each of the 20 isoaccepting groups.

Selection may not have favored a recognition system in which all 20 amino acids are specified by only acceptor stem nucleotides partly because this type of system is imperfect with respect to ensuring translational fidelity. In principle, translational fidelity could be achieved by an acceptor-stem-based recognition system as long as the nucleotide identity of the recognition element is not subject to change. However, mutation rates were probably very high in primordial systems due to errors introduced by imperfect replication systems. Mutations affecting the base pairs that comprise the acceptor stem would disrupt tRNA secondary structure and thus might be deleterious. More importantly, given the high degree of sequence overlap (Fig. 2), a mutation at the single-stranded nucleotide at position 73 would frequently change the amino acid specificity of the tRNA. For example, an A73G transition in the acceptor stem of a methionine tRNA would create a perfect asparagine tRNA acceptor stem sequence and would result in asparagine being inserted at methionine codons. Similarly, an A35G transition would change a valine anticodon (GAC) to alanine (GGC) and would cause valine to be inserted at alanine codons because the putative acceptor stem recognition site is intact. Thus an acceptor-stem-based recognition system would result in high levels of amino acid misincorporation. Moreover, if there had been a primordial recognition system based entirely on acceptor stem recognition, one would expect a suite of genetic codes rather than the nearly universal genetic code that is in evidence today.

Although acceptor stem nucleotides seem to have predominated in the earliest tRNA recognition systems, anticodon nucleotides are now more important to aminoacylation than are those in the acceptor stem. For example, *E. coli* glutaminyl-tRNA synthetase (GlnRS) specifically interacts with both an acceptor stem base pair (the absolutely conserved G2-C71) and all three anticodon nucleotides (Rould et al. 1989, 1991). Yet, anticodon nucleotides make larger contributions to aminoacylation kinetics than do the acceptor stem nucleotides (Jahn et al. 1991). Similar types of observations have been made for other isoaccepting groups (Pallanck and Schulman 1992; Giegé et al. 1993; McClain 1993; Saks et al. 1994). In all cases to date, when anticodon recognition has been found for an isoaccepting group in one organism, it has later been found to occur in the same isoaccepting group in other organisms. This supports the idea that anticodon recognition evolved prior to the split of the three major organismal groups and was an important and enduring innovation.

Anticodon recognition has clear advantages over acceptor stem recognition because it buffers translation with respect to mutations in tRNAs. In the extreme case, where tRNA amino acid identity is solely determined by anticodon nucleotides, mutations of these nucleotides would simultaneously change the tRNA amino acid identity and its mRNA coupling capacity (Engelhardt and Kisselev 1966). Mutations outside of the anticodon would be silent with respect to selection, unless they severely disrupted tRNA structure. Thus a system incorporating anticodon recognition would have a selective advantage over one based entirely on acceptor stem recognition and would help to stabilize the genetic code.

The availability of an anticodon stem/loop and the advantages of anticodon recognition may have driven the evolution of synthetases. After synthetases evolved a size sufficient to span the distance between the tRNA 3′-terminus and the anticodon they could have evolved a binding pocket having a specificity for the anticodon nucleotides of the cognate tRNAs. The structure of the synthetases could then be further refined such that tRNA (or anticodon) binding could contribute to catalysis via induced conformational changes in the tRNA, the synthetase, or both.

Structural Context of tRNA Evolution

It is likely that the evolution of a tRNA-like structure with an associated anticodon stem/loop was an important

innovation that provided a context for the refinement of tRNA recognition systems. Therefore, the tRNA sequences of the deep Archaea (above) were analyzed to explore when the tRNA-like structure evolved relative to the assignments of amino acids to anticodons. The tRNAs from these organisms were chosen because they provide the best available approximation of an early stage in tRNA evolution as well as a snapshot of current relationships among tRNA isoaccepting groups.

A total of 36 tRNA sequences of type I tRNAs representing 16 isoaccepting groups were included in the analysis. Sequences for the type II tRNAs (leucine and serine) were excluded because the extra nucleotides that form a stem in the variable loop confound distance calculations; and sequences for tryptophan and cysteine tRNAs from these organisms were not available. The *Methanobacterium thermoautotrophicum* $tRNA^{Gly}$ and $tRNA^{Ile}$ were included because only one of the possible three isoacceptors in each of these groups was available in the data set for the deep Archaea.

Preliminary analyses revealed that the tRNAs from the extremely thermophilic Crenarchaeota formed a series of deep branches within the tree and that the exclusion of these tRNAs from the data set did not affect the clusters formed among the other tRNAs. Consequently, tRNAs from the thermophilic Crenarchaeota (except $tRNA^{Met}$, because no other methionine tRNA sequence was available) were eliminated from the final sequence analysis to simplify the tree.

Homologous regions of the tRNAs were aligned according to the standard format that reflects tRNA secondary structural features (Gauss et al. 1979). We did not directly test our data base for phylogenetic signal (Hillis et al. 1993). Although there has been some concern that the divergence of tRNA sequences is so great that significant phylogenetic information has been lost (Holmquist et al. 1973), other workers have shown that significant information was still present in these molecules (Cedergren et al. 1980; Fitch and Upper 1987; Eigen et al. 1989). Prior to the analysis, the three anticodon nucleotides and the adjacent N37 were removed so that tRNA groupings would not be influenced by their decoding capabilities. Distances between sequences were calculated using the maximum-likelihood model and were analyzed using the Fitch-Margoliash method (Fitch and Margoliash 1967) in the Phylogenetic Inference Package (PHYLIP; Felsenstein 1993) with the input order of the species randomized ten times and without negative branch lengths. The Fitch-Margoliash method was used because it is slightly more accurate than is the neighbor-joining method when the substitution rates on different branches are unequal (Kuhner and Felsenstein 1994). The bootstrap scores obtained from 100 replicates of the data set (using the Boot and Consense programs of PHYLIP; Felsenstein 1993) are reported at the base of the nodes.

The tRNAs from the deep Archaea fell into four major clusters that comprise three groups of tRNAs having different sequence motifs (Fig. 3). Group C included glycine and valine tRNAs; group D included two clusters having proline, glutamine, histidine, asparagine, and glutamate tRNAs; and group A included the remaining tRNAs. The fairly low bootstrap scores at the nodes defining the four major clusters may be due to differences in the substitution rates at sites along the tRNA sequences. Differences in substitution rates at first, second, and third positions of the codons in mRNAs are known to exist. However, it is not yet clear how to categorize sites in structural RNAs with regard to the different substitution rates which surely also exist. Consequently, it was not possible to accommodate this type of variation in the analysis. In light of this inescapable weakness of the sequence analysis, it is quite remarkable that the clusters reflect three distinctive sequence motifs.

The three groups indicated in Fig. 3 reflect sequence motifs located within the core structure of the tRNAs. (See Fig. 1 for tertiary and triple-base pairs that dictate core structure.) The tRNAs in group A have the standard four base pairs in the D-stem and five nucleotides in the variable loop (D4/V5). These tRNAs probably have the overall tertiary structure that is illustrated by the crystal structure of yeast $tRNA^{Phe}$ (Kim et al. 1974; Ladner et al. 1975). Group B is a subset of group A containing *Methanococcus vannielii* $tRNA^{Phe}$ and $tRNA^{Tyr}$ and *M. voltae* $tRNA^{Tyr}$. These tRNAs differ from those in group A in having two transitions that result in a U11-A24 base pair (instead of a C11-G24) and a U45. In addition, they have a G10-C25-U45 triplet that may cause them to have a slightly different structure than group A tRNAs. The tRNAs in group C are also D4/V5 but differ from those in groups A and B by allowing a wobble-base pair at positions 10–25 and 13–22 and by having a Y13-G22-A46 triplet. Group D tRNAs differ from those in the other groups in having only four nucleotides in the variable loop and in always having a nucleotide at position 20a in the D-loop. In addition, they have an unconventional U13-U22 base pair, a predicted U13-U22-A46 triplet, and a standard Watson-Crick base pair between N26 and N44. By imagining that A21 might shift to the position occupied by U22 one obtains a tRNA having a standard U13-A22-A46 triplet and a nonstandard bulged U. Whether this structure is actually an isomorph of known tRNA structures remains to be seen. Nevertheless, tRNAs having this core structure form a unified group.

An association between core structures and anticodon nucleotides was examined to explore any possible correspondence between the evolution of the tRNA-like structure and the establishment of the amino acid–anticodon assignments that constitute the genetic code (Fig. 4). Most isoaccepting groups in group A have a U36 and those in group B have an A36. Similarly, isoaccepting groups in group C have a C36 and those in group D can be characterized as having either a U35 or a G36.

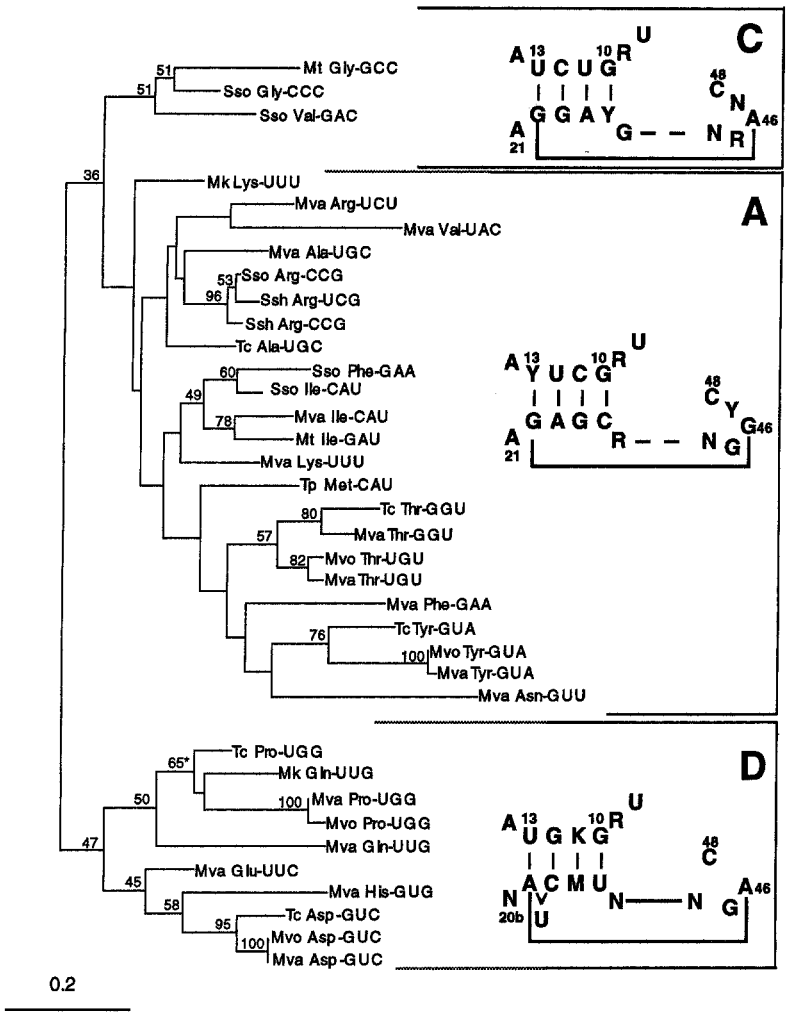

Fig. 3. A phylogenetic tree (unrooted) for 36 tRNA sequences from the deep Archaea. The figure presents the best tree obtained using the Fitch-Margoliash method (percent standard deviation = 20.6) without negative branch lengths. This tree is nearly identical to the majority-rule consensus tree with the one important exception noted below. Bootstrap scores obtained from 100 replicates are indicated at the base of the nodes defining the four major clusters and at nodes having scores greater than 45%. The *bar* below the tree indicates the expected number of changes accumulated for each site based on the average rate of change. The three groups indicated on the right side of the diagram reflect sequence motifs that are located within the core structures of each of the associated tRNAs (except Mva Arg-UCU, Mva Val-UAC, Tp Met-CAU, and Tc Tyr-GUA, whose core structures differ somewhat from those that are presented). The consensus sequence of these motifs is presented in the standard secondary structure format (Fig. 1). Nucleotide abbreviations are K = U or G; M = A or C; R = A or G; Y = U or C; N = any nucleotide. Organisms are designated by the following abbreviations: *Mva Methanococcus vannielii, Mvo Methanococcus voltae, Mk Methanopyrus kandleri, Sso Sulfolobus solfataricus, Ssh Sulfolobus shibatae, Tp Thermofilum pendens, Tc Thermococcus celer, Mt Methanobacterium thermoautotrophicum.* *This score was at a bifurcation in the consensus tree that resulted in one cluster containing Tc Pro-UGG and MK Gln-UUG and another cluster, as shown, containing Mva Pro-UGG and Mvo Pro-UGG.

The probability of finding the observed association between the core structure of each isoaccepting group and the middle (N35) or 3′ (N36) anticodon nucleotide was examined using Fisher's exact test (SAS 1992). Whereas core structures were independent of the nucleotide at position 35 ($P < 0.43$) there was a significant association between core structures and N36 ($P < 0.0005$). This association of structural features with anticodons is particularly striking because the anticodon nucleotides were eliminated from the tRNA sequences used in building the gene tree.

It is also intriguing that each of the three major structural groups contains an isoaccepting group having a C36 and that the four amino acids formed in greatest abundance by abiotic synthesis (alanine, glycine, valine, and aspartic acid; Miller 1987) are represented within this restricted subset. Taken together, the results support the idea that three subtly different tRNA-like structures were present close to the time that amino acid–anticodon assignments were first being established. These core structures may have provided three subtly different scaffolds for the presentation of recognition elements to the synthetases and may have thereby provided sufficient variation to promote the divergence of tRNAs into the 20 isoaccepting groups.

tRNA Gene Recruitment

An intriguing property of the tRNA gene tree (Fig. 3) is that members of some isoaccepting groups sometimes fail to cluster together; and when this occurs, they often cluster with tRNAs that differ by a single anticodon nucleotide. For example, one valine isoacceptor is clustered with glycine tRNAs whereas the other $tRNA^{Val}$ is clustered with arginine and alanine tRNAs. Similarly, one $tRNA^{Phe}$ is clustered with isoleucine tRNAs whereas the other $tRNA^{Phe}$ is clustered with tyrosine tRNAs. A cluster containing valine and glycine isoacceptors and another cluster containing valine and alanine isoacceptors was also found in our analysis of *E. coli* tRNAs (data not shown) and in an analysis of tRNA sequences that were drawn from several different organisms (Cedergren et al. 1980). Moreover, others have noted that tRNAs from different isoaccepting groups sometimes have very similar sequences (Squires and Carbon 1971; Staves et al. 1986).

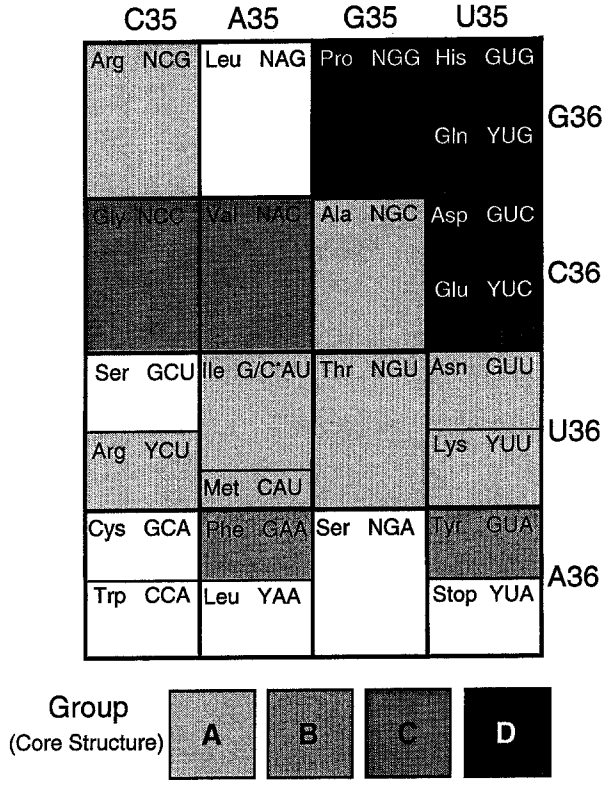

Fig. 4. Relationship between anticodons and tRNA core structures. The figure is arranged by anticodon nucleotide with N35 in *columns* and N36 in *rows*. The amino acid associated with each anticodon is given within each *box*. The arrangement differs somewhat from standard anticodon tables to facilitate showing the relationship between anticodon nucleotides and tRNA core structures (Fig. 3). *C34 is modified to lysidine.

To further investigate the implications of these observations with respect to the general problem of tRNA gene sequence evolution we analyzed 53 tRNA sequences from the extremely halophylic Archaea (*Halobacterium* and *Halococcus*). This group was chosen for two reasons. First, it is phylogenetically related to the deep Archaea (Woese et al. 1990; Olsen et al. 1994). Second, a substantial number of tRNA gene sequences, representing several isoacceptors within each tRNA family, were available.

The analysis of halophylic Archaea tRNA sequences was consistent with the analyses discussed above in that when tRNAs failed to faithfully cluster by their amino acid specificity, they often clustered with tRNAs that differed by a single anticodon nucleotide, even though these nucleotides were eliminated from the analysis. Differences in the locations of the arginine isoacceptors within the gene tree for the deep Archaea and halophylic Archaea were of particular interest. In the analysis of the deep Archaea, the arginine isoacceptors fell into group A (Fig. 3). However, in the analysis of the halophylic Archaea sequences, three arginine isoacceptors retained their affinity with group A tRNAs, whereas two others had moved into group D (data not shown). There are two possible explanations for how these arginine tRNAs acquired a group D structure: either they accumulated neutral mutations in their group A structure that ultimately moved them into group D, or they were recruited from a tRNA that was already in group D. The first explanation seems unlikely because it requires mutation rates to differ significantly among arginine tRNAs and because it requires mutations in nucleotides that dictate structure to be nondeleterious. In contrast, the results of our sequence analyses and the general observation that anticodon nucleotides can make important contributions to tRNA recognition (Pallanck and Schulman 1992; Saks et al. 1994) support the idea that arginine tRNAs could have been recruited into group D via an anticodon mutation in a tRNA from another isoaccepting group (such as Gln, Pro, or His) that already had a group D structure. We therefore propose that anticodon mutations can promote the recruitment of tRNAs from one isoaccepting group to another. Consequently, tRNA evolution may not always be monophyletic with respect to each isoaccepting group.

The overlapping acceptor stem sequences of tRNAs from different isoaccepting groups (cf. Table 1 and Fig. 2) could be important in predisposing groups of tRNAs to recruitment events when the cognate synthetase recognizes acceptor stem as well as anticodon nucleotides. In addition, there is some experimental evidence to support the idea that structural variations could also predispose tRNAs to recruitment events. A broad study of the in vivo amino acid identity of amber suppressors (anticodon CUA) showed that amber suppressors derived from five different isoaccepting groups were misacylated by GlnRS (Normanly et al. 1990). The suppressor derived from the GlyT isoacceptor was among the tRNAs that were misacylated by GlnRS whereas the suppressor derived from the GlyU isoacceptor retained its original amino acid identity. These two glycine tRNAs have nearly identical acceptor stem sequences and do not have either the weak 1-72 base pair or the G2-C71 base pair that would predispose them to recognition by GlnRS (Rould et al. 1989, 1991; Jahn et al. 1991). However, they differ in structural features. GlyT falls within the group D structural class, as do the deep archaeal glutamine tRNAs, whereas the GlyU isoacceptor does not. Thus it is possible that the structural features of GlyT predisposed it to misacylation by GlnRS once the anticodon was changed.

Because recruitment involves a loss as well as a gain of function it could only occur under certain circumstances. The genome must have multiple copies of the tRNA whose function will be lost when recruitment occurs or it must have a set of isoacceptors with overlapping mRNA coupling capacities. Otherwise a recruitment event would be lethal due to the loss of a tRNA critical for translation. Moreover, because recruitment is only likely to be favored when a wild-type copy of a tRNA has been debilitated by mutation, a recruited tRNA should have an activity that at least approximates that of a debilitated wild-type tRNA and that can be improved

rapidly by subsequent mutations. In particular, when the recruited tRNA is expressed at normal levels, it would have to be aminoacylated at a rate sufficient to sustain normal translation, interact correctly with a particular codon in all types of endogenous mRNAs, work efficiently on the ribosome, and not have a mixed amino acid identity, because this would be lethal.

The results of some in vitro (Schulman and Pelka 1988, 1989, 1990) and in vivo studies (McClain et al. 1990; Normanly et al. 1990; Pallanck and Schulman 1991), in conjunction with the results of our sequence analyses, are consistent with the recruitment hypothesis. However, although previous experiments revealed effects of anticodon nucleotides on tRNA amino acid identity, they were not designed to directly test the recruitment hypothesis. We are currently testing this hypothesis by knocking out an essential tRNA gene in *E. coli* and by determining whether a tRNA that we recruited via an anticodon mutation can compensate for the loss of the wild-type activity. Our preliminary results indicate that we have successfully recruited a $tRNA^{Arg}$ to compensate for the loss of an essential $tRNA^{Thr}$.

Conclusions

Transfer RNA sequences have been used to test hypotheses about the order in which amino acids were assigned to the genetic code, and the age of the code, and to deduce the possible nucleotide sequence of the progenitor tRNA (Cedergren et al. 1980; Fitch and Upper 1987; Eigen et al. 1989; Rodin et al. 1993; Szathmáry 1993). We have used tRNA sequences to explore the nature of primordial recognition systems, how these systems may have influenced the evolution of the genetic code, and the impact of different types of recognition systems on tRNA gene evolution.

The evolution of the genetic code was probably a slow process that was dependent on the appearance of innovations that improved the specificity of tRNA-synthetase interactions. Because evolution is a tinkerer (Jacob 1982), important relicts of old recognition systems must have been retained as new innovations were incorporated. Recent evidence that aminoacylation can be accomplished by an RNA catalyst (Illangasekare et al. 1995) suggests that the first synthetases were RNAs rather than proteins. However, there were probably only a few types of isoaccepting groups during the time when tRNAs were recognized by putative primordial RNA synthetases because the specificity of these synthetases was probably poor and because no evidence remains of the involvement of an RNA cofactor (except ATP) in the aminoacylation reaction. In contrast, the spliceosome, RNase P, telomerase, and probably peptidyl transferase all have retained their integral RNA component (cf. Gestland and Atkins 1993). Therefore, the expansion of the genetic code probably required peptides that assisted RNAs in the aminoacylation reaction and that replaced the RNA synthetases before all of the codon assignments were made.

In all but four tRNA families, at least one acceptor stem base pair or the nucleotide at position 73 is conserved (Table 1). Interestingly, there is a relationship between synthetase class (Eriani et al. 1990) and the nucleotide that is conserved at position 73. Eight of the ten tRNAs that are aminoacylated by class I synthetases have an A73 (the exceptions being $tRNA^{Arg}$ and $tRNA^{Cys}$) whereas those that are aminoacylated by class II synthetases have greater nucleotide diversity at position 73. The acceptor stem sequence conservation, in conjunction with biochemical results showing that acceptor stem nucleotides can contribute to recognition of the tRNAs of extant organisms (Schimmel et al. 1993), is consistent with the idea that acceptor stem recognition is an ancestral trait. However, because of the overlap among the acceptor stem sequences of tRNAs from different isoaccepting groups (Fig. 2), it seems that the 20 isoaccepting groups were not completely defined by nucleotides in this region of the molecule.

Acceptor stem recognition may have begun as a means for aminoacylating small RNA hairpins and was probably retained after the tRNA structure evolved. However, the difficulties inherent in the recognition of base-paired RNA by either an RNA or a protein synthetase probably restricted the number of amino acids that could be faithfully specified by acceptor stem recognition. This bottleneck may have been overcome with the advent of a tRNA-like structure having an anticodon stem/loop and by subsequent changes in the size and sophistication of the synthetases. Once synthetases could span the approximately 70 Å between the tRNA 3′-terminus and the anticodon they could specifically interact with not only acceptor stem but also anticodon nucleotides. That the advent of anticodon recognition helped to accurately specifying all 20 isoaccepting groups fits well with Jukes's ideas about the expansion of the genetic code and his codon capture hypotheses (cf. Osawa et al. 1992). Once synthetases evolved the capacity to recognize anticodon nucleotides, selection would have favored this system relative to the misincorporation-prone acceptor stem recognition system. Thus the genetic code may have evolved through a series of steps that were dictated by the availability of different types of tRNA recognition systems.

Strong recognition of the anticodon by synthetases endows tRNA gene sequence evolution with interesting and unusual properties. Transfer RNA gene sequences can be quite evolutionarily labile without sacrificing the high degree of translational fidelity that is necessary in the cell. Neutral mutations will accumulate in regions of the tRNA outside of the anticodon. In addition, strong anticodon recognition will tend to intermix tRNA gene sequences due to recruitment events caused by anticodon mutations. In addition, strong anticodon recognition gen-

erates the unusual property that mutations in a functionally important part of the tRNA molecule can be selectively neutral (nondeleterious).

Acknowledgments. We are grateful to John Abelson, in whose laboratory this work was performed, for his insights about tRNA recognition systems and his support of our forays, past and present, into tRNA sequence analyses. We thank David Mathog, Michael Alvarez, and Jim Archie for their assistance with computations, Ron Swanson for providing tRNA sequences from *Pyrobaculum,* and Hyman Hartman for organizing a stimulating conference on "The Aminoacyl-tRNA Synthetases and the Evolution of the Genetic Code." We are also grateful to Andrea Ghetti, Jody Hey, and Tony Zera, who provided detailed comments that led to significant improvements in the manuscript. This work was supported by NIH grant GM 48560.

References

Cedergren RJ, LaRue B, Sankoff D, Lapalme G, Grosjean H (1980) Convergence and minimal mutation criteria for evaluating early events in tRNA evolution. Proc Natl Acad Sci USA 77:2791–2795

Eigen M, Lindemann BF, Tietze M, Winkler-Oswatitsch R, Dress A, von-Haeseler A (1989) How old is the genetic code? Statistical geometry of tRNA provides an answer. Science 244:673–679

Engelhardt VA, Kisselev LL (1966) Recognition Problem: On the specific interaction between coding enzyme and transfer RNA. In: Kaplan NO, Kennedy EP (eds) Current aspects of biochemical energetics. Academic Press, New York pp 213–225

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347:203–206

Felsenstein J (1993) PHYLIP (phylogeny inference package). Department of Genetics, University of Washington, Seattle. Distributed by the author

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Fitch WM, Upper K (1987) The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. Cold Spring Harbor Symp Quant Biol 52:759–767

Francklyn C, Schimmel P (1990) Enzymatic aminoacylation of an eight-base-pair microhelix with histidine. Proc Natl Acad Sci USA 87:8655–8659

Gauss DH, Grüter F, Sprinzl M (1979) Proposed numbering system of nucleotides in tRNAs based on yeast $tRNA^{Phe}$. In: Schimmel PR, Söll D, Abelson JN (eds) Transfer RNA: structure, properties, and recognition. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY pp 518–519

Gestland RF, Atkins JF (eds) (1993) The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Giegé R, Puglisi JD, Florentz C (1993) tRNA structure and aminoacylation efficiency. Prog Nucleic Acid Res Mol Biol 45:129–206

Hillis DM, Allard MW, Miyamoto MM (1993) Analysis of DNA sequence data: phylogenetic inference. Methods Enzymol 224:456–487

Himeno H, Hasegawa T, Ueda T, Watanabe K, Miura K-I, Shimizu M (1989) Role of the extra G-C pair at the end of the acceptor stem of $tRNA^{His}$ in aminoacylation. Nucleic Acids Res 17:7855–7863

Holmquist R, Jukes TH, Pangburn S (1973) Evolution of transfer RNA. J Mol Biol 78:91–116

Hou Y-M, Schimmel P (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. Nature 333:140–145

Illangasekare M, Sanchez G, Nickles T, Yarus M (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. Science 267:643–647

Iwabe N, Kuma K-I, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 86:9355–9359

Jacob F (1982) The possible and the actual. University of Washington Press, Seattle

Jahn M, Rogers MJ, Söll D (1991) Anticodon and acceptor stem nucleotides in $tRNA^{Gln}$ are major recognition elements for *E. coli* glutaminyl-tRNA synthetase. Nature 352:258–260

Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, Wang AHJ, Seeman NC, Rich A (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science 185:435–440

Kuhner MK, Felsenstein J (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11:459–468

Ladner JE, Jack A, Robertus JD, Brown RS, Rhodes D, Clark BFC, Klug A (1975) Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. Proc Natl Acad Sci USA 72:4414–4418

Maizels N, Weiner A (1994) Phylogeny from function: Evidence from the molecular fossil record that tRNA originated in replication, not translation. Proc. Natl. Acad Sci USA 91:6729–6734

McClain WH (1993) Rules that govern tRNA identity in protein synthesis. J Mol Biol 234:257–280

McClain WH, Foss K (1988) Changing the identity of a tRNA by introducing a G-U wobble pair near the 3′ acceptor end. Science 240:793–796

McClain WH, Foss K, Jenkins RA, Schneider J (1990) Nucleotides that determine *Escherichia coli* $tRNA^{Arg}$ and $tRNA^{Lys}$ acceptor identities revealed by analyses of mutant opal and amber suppressor tRNAs. Proc Natl Acad Sci USA 87:9260–9264

Miller SL (1987) Which organic compounds could have occurred on the prebiotic earth. Cold Spring Harbor Symp Quant Biol 52:17–27

Möller W, Janssen GMC (1990) Transfer RNAs for primordial amino acids contain remnants of a primitive code at position 3 to 5. Biochimie 72:361–368

Möller W, Janssen GMC (1992) Statistical evidence for remnants of the primordial code in the acceptor stem of prokaryotic transfer RNA. J Mol Evol 34:471–477

Normanly J, Kleina LG, Masson J-M, Abelson J, Miller JH (1990) Construction of *Escherichia coli* amber suppressor tRNA genes III. Determination of tRNA specificity. J Mol Biol 213:719–726

Olsen GJ, Woese CR, Overbeek R (1994) The winds of (evolutionary) change: breathing new life into microbiology. J Bacteriol 176:1–6

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264

Pallanck L, Schulman LH (1991) Anticodon-dependent aminoacylation of a noncognate tRNA with isoleucine, valine and phenylalanine *in vivo*. Proc Natl Acad Sci USA 88:3872–3876

Pallanck L, Schulman LH (1992) tRNA discrimination in aminoacylation. In: Hatfield DL, Lee BJ, Pirtle R (eds) Transfer RNA in protein synthesis. CRC Press, Boca Raton, FL, pp 279–318

Rodin S, Ohno S, Rodin A (1993) Transfer RNAs with complementary anticodons: Could they reflect early evolution of discriminative genetic code adaptors? Proc Natl Acad Sci USA 90:4723–4727

Rould MA, Perona JJ, Söll D, Steitz TA (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. Science 246:1135–1142

Rould MA, Perona JJ, Steitz TA (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. Nature 352:213–218

Saks ME, Sampson JR, Abelson JN (1994) The transfer RNA identity problem: a search for rules. Science 263:191–197

SAS Institute Inc (1992) Frequency procedure. In: SAS procedures guide, version 6. SAS Institute Inc. Cary, NC, pp 331–340

Schimmel P, Giegé R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. Proc Natl Acad Sci USA 90:8763–8768

Schulman LH, Pelka H (1988) Anticodon switching changes the identity of methionine and valine transfer RNAs. Science 242:765–768

Schulman LH, Pelka H (1989) The anticodon contains a major element of the identity of arginine transfer RNAs. Science 246:1595–1597

Schulman LH, Pelka H (1990) An anticodon change switches the iden-

tity of *E. coli* $tRNA^{Met}m$ from methionine to threonine. Nucleic Acids Res 18:285–289

Squires C, Carbon J (1971) Normal and mutant glycine transfer RNAs. Nature New Biol 233:274–277

Staves MP, Bloch DP, Lacey JC (1986) Evolution of *E. coli* $tRNA^{Ile}$: evidence of derivation from other tRNAs. Z Naturforsch 42c:129–133

Steinberg S, Misch A, Sprinzl M (1993) Compilation of transfer-RNA sequences and sequences of transfer-RNA genes. Nucleic Acids Res 21:3011–3015

Szathmáry E (1993) Coding coenzyme handles: a hypothesis for the origin of the genetic code. Proc Natl Acad Sci USA 90:9916–9920

Weiner AM, Maizels N (1987) Transfer-RNA like structures tag the 3′ ends of genomic RNA molecules for replication: implications for the origin of protein-synthesis. Proc Natl Acad Sci USA 84:7383–7387

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576–4579